DOCTOR OF PHILOSOPHY

On Statistical Analysis of Vehicle Time-Headways Using Mixed Distribution Models

Yu, Fu

Award date:
2014

On Statistical Analysis of Vehicle Time-Headways Using Mixed Distribution Models

Fu Yu

2014

University of Dundee

UNIVERSITY OF DUNDEE

On Statistical Analysis of Vehicle Time-Headways Using Mixed Distribution Models

Fu Yu

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

College of Art, Science & Engineering
School of Engineering, Physics & Mathematics

August 2014

Declaration of Authorship

I, Fu Yu, declare that this thesis titled, 'On Statistical Analysis of Vehicle Time-Headways Using Mixed Distribution Models' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ______________________________________

Date: ______________________________________

UNIVERSITY OF DUNDEE

Abstract

College of Art, Science & Engineering
School of Engineering, Physics & Mathematics

Doctor of Philosophy

Fu Yu

For decades, vehicle time-headway distribution models have been studied by many researchers and traffic engineers. A good time-headway model can be beneficial to traffic studies and management in many aspects; e.g. with a better understanding of road traffic patterns and road user behaviour, the researchers or engineers can give better estimations and predictions under certain road traffic conditions and hence make better decisions on traffic management and control. The models also help us to implement high-quality microscopic traffic simulation studies to seek good solutions to traffic problems with minimal interruption of the real traffic environment and minimum costs.

Compared within previously studied models, the mixed (SPM and GQM) models, especially using the *gamma* or *lognormal* distributions to describe followers headways, are probably the most recognized ones by researchers in statistical studies of headway data. These mixed models are reported with good fitting results indicated by goodness-of-fit tests, and some of them are better than others in computational costs. The gamma-SPM and gamma-GQM models are often reported to have similar fitting qualities, and they often out-perform the lognormal-GQM model in terms of computational costs. A lognormal-SPM model cannot be formed analytically as no explicit Laplace transform is available with the *lognormal* distribution. The major downsides of using mixed models are the difficulties and more flexibilities in fitting process as they have more parameters than those single models, and this sometimes leads to unsuccessful fitting or unreasonable fitted parameters despite their success in passing GoF tests. Furthermore, it is difficult to know the connections between model parameters and realistic traffic situations or environments, and these parameters have to be estimated using headway samples.

Hence, it is almost impossible to explain any traffic phenomena with the parameters of a model. Moreover, with the *gamma* distribution as the only common well-known followers headway model, it is hard to justify whether it has described the headway process appropriately. This creates a barrier for better understanding the process of how drivers would follow their preceding vehicles.

This study firstly proposes a framework developed using MATLAB, which would help researchers in quick implementations of any headway distributions of interest. This framework uses common methods to manage and prepare headway samples to meet those requirements in data analysis. It also provides common structures and methods on implementing existing or new models, fitting models, testing their performance hence reporting results. This will simplify the development work involved in headway analysis, avoid unnecessary repetitions of work done by others and provide results in formats that are more comparable with those reported by others.

Secondly, this study focuses on the implementation of existing mixed models, i.e. the gamma-SPM, gamma-GQM and lognormal-GQM, using the proposed framework. The lognormal-SPM is also tested for the first time, with the recently developed approximation method of Laplace transform available for *lognormal* distributions. The parameters of these mixed models are specially discussed, as means of restrictions to simplify the fitting process of these models. Three ways of parameter pre-determinations are attempted over gamma-SPM and gamma-GQM models.

A couple of response-time (RT) distributions are focused on in the later part of this study. Two RT models, i.e. Ex-Gaussian (EMG) and inverse Gaussian (IVG) are used, for first time, as single models to describe headway data. The fitting performances are greatly comparable to the best known *lognormal* single model. Further extending this work, these two models are tested as followers headway distributions in both SPM and GQM mixed models. The test results have shown excellent fitting performance. These now bring researchers more alternatives to use mixed models in headway analysis, and this will help to compare the behaviours of different models when they are used to describe followers headway data. Again, similar parameter restrictions are attempted for these new mixed models, and the results show well-acceptable performance, and also corrections on some unreasonable fittings caused by the over flexibilities using 4- or 5- parameter models.

Acknowledgements

Completing my PhD degree is probably the most challenging activity of my first 37 years of my life. The best and worst moments of my doctoral journey have been shared with many people. It has been a great privilege to spend several years in the School of Engineering, Physiscs & Mathematics at University of Dundee, and its members will always remain dear to me.

First and foremost I wish to thank my supervisor Professor Allan Gillespie. It has always been a great pleasure to work with him in all the past years. Without his kindness advice and support, it would not have been possible for me to accomplish all the work involved. He has been always generously sharing his experience not only in academic supports but also in many aspects of personal life. To me, he is not just my supervisor or a colleague, but he is also my great friend. His joy and enthusiasm towards his work has deeply influenced me, and I believe these precious attitudes will be very beneficial to me in my future work and life.

A special thanks to Professor Eddie Wilson (university of Bristol) , for the inspiring discussions in traffic flow theory and platoon formulations, which provided me great help and encouragement. During a very difficult mid stage of my study, these discussions enlightened the route of my studies. Also I'm very appreciating of the headway data sets provided by him, which saved me tremendous efforts. Professor John Nelson (University of Aberdeen) introduced me to the Universities Transport Study Group (UTSG), which provided me opportunities to know more members at its conferences. I gained much help from the community, by listening to the talks and having discussions with individuals, and I'd like to thank you all for the kindness and help you have offered. Many of my friends, Professor Pin Lin, Dr Kaixin Zhou etc., have supported me in many aspects via discussions, general researches, statistical modelling and analysis etc. These have given much encouragement and confidence.

I would also like give my thanks to my colleagues in our research group Material and Photonic System (MAPS). I'm very glad to have you all here for these years, and we have shared great working experiences, environments and friendships. A special thanks to my colleague and friend, Dr Guang Tang, for many useful discussions mainly in work involved in modelling and mathematics. Also Dalong Zhang, thank you for all discussions we had together in the fields of transport and ITS in the early stage. To my colleagues who are academic members in the Division of Physics, Professor Mervyn Rose, Dr Steve Reynolds, Dr Kazem Dastoori, Dr Yongchang Fan and Professor Amin Abdolvand etc. You all have been greatly supportive and offering generous helps whenever I need it. Also I'd like to have a big thanks to our supportive technicians, Dr Gary Callon, Grant Kydd, David Husband etc., and our administrative staff, Linda Rannie, Prue Reid, Karen Wilson etc.

I wish to thank my beloved parents, for their endless love, support and encouragement. Their love provided my inspiration and was my driving force. I owe them everything and wish I could show them just how deeply my love and appreciations to them. My dear wife, Weiwei Lee, whose love and support allowed me to accomplish this journey. Thank you for being here with me at all the difficult moments of my life, and thank you for all the great support to myself and our family. Finally a great thanks to all my friends, relatives and to people who are helping and caring about me, I hope that this work makes you all proud.

Fu Yu
Dundee University
July 2014

Contents

List of Figures

List of Tables

Abbreviations

AD test	Anderson-Darling goodness-of-fit test
BRT	brake reaction time
CDF	Cumulative distribution function
CV	Coefficient of variance
EMG	Exponentionally modified Gaussian
	(often known as ex-Gaussian distribution)
EMG-GQM	GQM, using EMG followers headway distribution
EMG-SPM	SPM, using EMG followers headway distribution
Eq.	Equation
gamma-GQM	GQM, using gamma followers headway distribution
gamma-SPM	SPM, using gamma followers headway distribution
GoF test	Goodness-of-fit test
GQM	Generalized queuing model
GUI	Graphical user interface
ITS	Intelligent transport system
IVG	Inverse Gaussian distribution
IVG-GQM	GQM, using IVG followers headway distribution
IVG-SPM	SPM, using IVG followers headway distribution
KS test	Kolmogorov-Smirnov goodness-of-fit test
lognormal-GQM	GQM, using lognormal followers headway distribution
lognormal-SPM	SPM, using lognormal followers headway distribution
M4	Cowan's mixed headway model, is the same as GQM
MCS	Minimum of chi-squared statistics
ME	Moment estimator

MISE	Minimum of the mean integrated squared errors
MLE	Maximum likelihood estimator
MMLE	Modified maximum likelihood estimator
MSSE	Minimum of the sum of the squared errors
RT model	Response time model
SPM	Semi-Poisson model
TH	Time headway

Notations

α	Shape parameter of the gamma distribution
β	Scale parameter of the gamma distribution
$\gamma(a, b)$	Incomplete gamma function
δ	Location parameter in shift-single models
λ	Arrival rates parameter in SPM and GQM mixed models
θ	Parameters of a distribution
$\hat{\theta}$	Maximum likelihood estimator for θ
μ	Parameter used in lognormal, EMG, IVG
σ	Parameter used in lognormal, EMG, IVG
τ	Parameter used in EMG
ϕ	Propotion parameter in SPM and GQM mixed models
$\mathcal{L}$	Likelihood function
q	Traffic flow rate, veh/hour
$f(t)$	Probablity density function (PDF)
$F(t)$	Cumulative distribution function
$g(t)$	PDF of single models or followers headway distributions
$g_e(t)$	PDF of EMG distribution used in mixed models
$g_g(t)$	PDF of gamma distribution used in mixed models
$g_i(t)$	PDF of IVG distribution used in mixed models
$g_l(t)$	PDF of lognormal distribution used in mixed models
$g^*(s)$	Laplace transform of $g(t)$ function
$G(t)$	CDF of single models or followers headway distributions

$G_e(t)$	CDF of EMG distribution used in mixed models
$G_g(t)$	CDF of gamma distribution used in mixed models
$G_i(t)$	CDF of IVG distribution used in mixed models
$G_l(t)$	CDF of lognormal distribution used in mixed models
$h(t)$	PDF of non-followers headway, exponential distribution
$H(t)$	CDF of non-followers headway, exponential distribution
T	Followers headway threshold
X	Time-headway random variable
U	Followers headway random variable
V	Non-followers headway random variable
$W(x)$	Lambert W function

Dedicated to my beloved parents,
For their endless love, support and encouragement...

Chapter 1

Introduction

Summary

A brief definition and introduction are given to explain vehicle time headways, together with the importance of headway in traffic-flow related areas. Some of the previous important research work is discussed. The objectives of this study are explained, and followed with a summarized structure of this thesis.

1.1 Introduction to Vehicle Time Headways

Vehicle time headway (TH), also named as headway for short, is the time interval measure between two consecutive vehicles (front bumper to front bumper) passing the same designated test point. A headway (measured in seconds) consists of two components, it is the sum of the time that the front vehicle takes to pass the test point and the time gap between two vehicles. In a quiet road, where the vehicle behind is not closely following the front one, the headways is presenting as random arrival time to the test point, as no delay has been caused by queuing. However, in a queue of traffic, the first component is depending on the length of the front vehicle and the velocity it is travelling. The second component is often considered as the safety gap that the following vehicle will maintain for safe driving.

The time intervals between vehicles are more of interest to individual drivers for safety reasons, while headways are more important to traffic flow theory in research and traffic engineering areas. Headways have a very close relationship to traffic flows, which measure the number of vehicles passing the same road point per unit time, e.g. vehicles per hour (veh/hr). Since the mean of headways is the reciprocal of traffic flow rate, vehicle headways represent microscopic measures of flows passing a point.

Headway is important as, in the first instance, it can be easily collected by many modern technologies, such as inductance detectors, varieties of traffic counters or video cameras. To some extent, headways can be looked upon as the footprint of drivers following behaviours. On the one hand, mean headway reflects the traffic levels overall. On the other hand, headways reflect how drivers are following their front vehicles if travelling within a queue.

1.2 Importance of Time Headway Study

Accurate modelling and analysis of headway distribution helps traffic engineers to maximize roadway capacity and minimize vehicle delays. Highway capacity is linked with headway distributions (Buckley, 1968; Minderhoud et al., 1997), as the capacity can be measured, in one way, using the mean of followers headway in platoons. The 1985 Highway Capacity Manual contains specification for computing capacities and levels of service for un-signalized intersections (Transportation Research Board, 1985). Headway studies are also involved as a crucial role in analysis of un-signalized intersections and roundabouts (Sullivan and Troutbeck, 1994).

Headway studies are closely related to vehicle merging, lane changing and turning operations. A number of microscopic simulation models have been developed for various road configurations. Knowledge of headways play a key role in generating vehicle arrival times as input into the simulation process. In fact, many of the researches on headway distribution analysis are motivated by the needs of developing simulation models. In addition, lane changing models or traffic merging models are highly dependent on the justification of availability of gaps, which is again dependent on headway distributions. Much research covers specific studies on gap availability (May, 1965) and acceptance (Drew, 1968) or ramp merging models (Wohl and Martin, 1967).

In interrupted traffic applications, gap acceptance studies are important for signalized and sign-controlled intersections (Blunden et al., 1962; McDowell et al., 1983), where headway could be used to minimize time delay by synchronizing with signal plans (Zhang et al., 2007).

In recent years, with the rapid development of information and communication technologies, the intelligent transport system (ITS) plays a more important role in many respects on improving traffic congestions and safety issues by providing better information and management to road users (Chowdhury and Sadek, 2003; Figueiredo et al., 2001). With the demand of ITS development, understanding of the headway data, to better understand drivers traffic behaviours, is essential.

All the above indicate that headway studies are fundamental in addressing many traffic engineering, traffic flow research and simulation issues.

1.3 Statistical Studies of Headway Models

Traffic flows have often been considered as stochastic processes since 1930s (Adams, 1936), headways are often studied using distribution models in a statistical analysis. In the past decades, many sophisticated models have been proposed or developed and many of these have been carefully reviewed and examined by researchers (Ha et al., 2012; Hoogendoorn and Botma, 1996; Hoogendoorn and Bovy, 1998; Luttinen, 1996; May, 1990; Zhang et al., 2007).

Some characteristics of headways are widely described and presented using the mean, median, mode and coefficient of variation (Breiman and Lawrence, 1977; Buckley, 1968; Dunne et al., 1968; Griffiths and Hunt, 1991; May, 1961, 1965, 1990). More generally, most headway studies use headway distribution models and histograms to describe the empirical headway data.

Some categorizations of headway distribution models have been suggested. A three-category suggestion used by Ha et al. (2012) seems to be reasonably precise. He justified the importance of followers behaviour, and classified the models into single, combined and mixed models. Others have used a two-category classification (Luttinen, 1996; Zhang et al., 2007), with only single and mixed models.

Single models

Single models, also known as simple models, have relatively simple forms of formulations, and they mainly use the standard probability distributions to model the headway variables. The common widely used models are *negative exponential, gamma* and *lognormal* distributions, etc. In single models, all vehicles are considered to have only one behaviour. The first model proposed by Adams (1936) is the *exponential* distribution. It assumes that vehicle arrivals are drawn from a Poisson process. The *exponential* model has its good theoretical foundation hence it has been referred to by many later studies. However, the model has very limited fitting performance for empirical headway data. It is more considered as the random process of the arrivals for platoons (or random bunches) (Branston, 1976; Buckley, 1968; Miller, 1961).

The *lognormal* distribution model (Daou, 1964; Greenberg, 1966) has gained probably the most attention as a single model, because of its close connection with the microscopic traffic flow theory, and more specifically the car-following models. Daou (1966) has presented detailed analysis of headway data and theoretical justifications of using the *lognormal* model for constrained headways.

The *gamma* distribution is often used in headway analysis due to its flexibility and compatibility, as it is the generalization of the *exponential* distribution and it has connections with many other distributions. The *gamma* distribuition can be evolved to *exponential* or *chi-square* distributions when one of the two parameters is fixed at specific values (Zhang et al., 2007). The *gamma* model can also be generalized to the *Pearson type III* model by introducing a location parameter. The *Erlang* distribution can be derived by using only integers as values of the shape parameter of *gamma* distribution. May (1990) suggested that this family of models is more suitable for describing headways at intermediate traffic flow levels.

Single models are simple and easy to apply, however, they generally have very limited performance especially in approximating short or very short headways (Ha et al., 2012). To improve their accuracy, a location parameter δ is often applied to form shifted-single models. Studies show some shifted-single models have significant improvement on their performances. Similar tests are also attempted in the latter part of chapter 2 in this thesis, for the *gamma* and *lognormal* distributions as two examples of single models.

Compared with single models that consider all vehicles with one single behaviour, most of the combined and mixed models consider the headway distribution as a combination or mixture of two group of drivers, i.e. followers and non-followers (leaders of queues, or free-driving vehicles). Hence researchers have developed models with components of two distributions. The proportion factor ϕ is used in both cases, to describe the proportion of followers in all vehicles.

Combined models

The combined models use a headway threshold to distinguish the followers and non-followers groups. These models use a simple linear addition of two models, and one is always an *exponential* distribution to describe headways of non-followers. The followers are considered to be drawn from one of the single models, such as *gamma*, *lognormal*, *Erlang* etc. Then the new models are named as hyper-single models, such as *Hypergamma*, *Hyperlognormal*, *Hyperlang* etc.

Some distributions used as followers headway models have been studied widely in the past. Dawson and Chimini (1968); Drew (1967) proposed the *Hyperlang* model, which had been further generalized into the *Hypergamma* model (Ha et al., 2011). The *normal* distribution was suggested by May (1990) as a followers headway model, however, this approach is not well accepted due to its poor fitting performance. Ha et al. (2012) tested four combined models in his study, and he reported that *Hyperlognormal* is the best performing model compared with the others.

Another combined model of the *Double Displaced Negative Exponential* Distribution (DDNED) is proposed by Griffiths and Hunt (1991), who reported good fitting results to headway data. Zhang et al. (2007) also tested the DDNED model in his study, and he found its performance outstanding.

Some studies (Ovuworie et al., 1980) have proposed models with more than two components, however, these studies are not widely applied due to their overly complex formulation and implementation.

Generally, combined models have some improved performance compared with single models, with the use of an exponential distribution of non-followers group. However, the implementation results reported are still not well accepted in terms of their fitting performance, plus, applying arbitrary threshold to separate followers and non-follower groups are considered to lack theoretical background. For these reasons, combined models are not within the scope of the models of attention in this study.

Mixed models

Instead of using a headway threshold to categorize the groups of followers and non-followers, the mixed models assume that the two groups have a statistical relationship. Two types of mixed models have been suggested in the past (Branston, 1976; Buckley, 1968; Cowan, 1975), i.e. the semi-Poisson model (SPM) and the generalized queuing model (GQM). SPM models use the followers headway distribution to formulate a conditional probability model for non-followers, which suppose the non-followers would have larger random headway variables than the followers, and the headway variables are drawn from an *exponential* distribution. GQM models assume that every non-followers headway consists of the sum of two parts, i.e., the exponential part plus a service time which is the same as the followers headway distribution. Studies show that both SPM and GQM are similar in the result of parameter estimation and goodness-of-fit test (Ha et al., 2012; Luttinen, 1996). Some works(Luttinen, 1996; Wasielewski, 1979) have concerns that GQM is lacking in physical basis, while others (Branston, 1976; Ha et al.,

2012) give GQM more credibility in terms of better accuracy and simplicity in implementation.

Followers headway models using *gamma, lognormal, normal* and *Rayleigh* distributions have been proposed and examined in studies. Overall, the fitting of the gamma-SPM (Buckley, 1968; Luttinen, 1996), gamma-GQM (Ha et al., 2012) and lognormal-GQM (Branston, 1979) have been recommended with outstanding fitting performance in different aspects. Many attempts have been made using mixed models to have better understanding of car following behaviours. Nowadays, SPM and GQM models are probably recognized as the most attractive models that have drawn much attention. In the present study, some existing and newly proposed mixed models are implemented and tested as one of the major objectives.

Chapter 4 contains further discussion between combined and mixed models. Luttinen (1996) reviewed comprehensively many of the above models, and discussed each model in terms of three criteria: *reasonability, applicability* and *validity*. Luttinen concluded with a recommendation of the gamma-SPM, because of its outstanding performance in various aspects. Zhang et al. (2007) compared a selection of single and mixed models and he found that the *DDNED* model out-performed other models. Ha et al. (2012) also made a comprehensive comparison of models, and he found that the gamma-GQM achieved the best performance in his study.

1.4 Problems in Headway Statistical Analysis

The first practical difficulty in headway analysis is probably the issues in comparing all the work reported by others. This is mainly due to the lack of common procedures in dealing with headway data, models and results. These are manifested in the following ways:

1. The presentations of headway data are often very different, and it is hard to categorise and compare the data among the various studies;
2. The sampling methods used in data preparation are often very different;
3. The fitting methods that work in one study may not work in others;
4. The goodness-of-fit (GoF) tests used are often different;
5. The presentations of results are often very different.

Luttinen (1996) pointed out some similar issues in data processing and GoF tests. As a result, Luttinen provided comprehensive suggestions in testing and comparison of models, especially in applying more powerful GoF test methods.

The second problem concerns variations and uncertainties of parameter estimations in the stage of fitting models to sampled headway data. As described in chapters 2 and 4, many types of estimation methods have been attempted in fitting mixed models. Due to their complexity in form they have more parameters, these estimations often have to be solved numerically through optimization, and the fitting results are sometimes sensitive to initial values. This could lead to failures or unreasonable results of estimations. The blame for such faults is often ascribed to either bad sampling of data or complications of models. This is often a corollary of the third problem, that is, researchers have paid more attention to the goodness of fit for their result, but less to the variations of parameters of the models. Many studies have used various GoF test methods to show how good models are capable of fitting data, but often they have ignored any implicit connections between the parameters of a 'good' model and other potential factors that may be relevant.

Some researchers have investigated parameters, and revealed potential connections between them and traffic conditions. For instance, Wasielewski (1979) interpreted the proportion parameter and arrival rates using traffic flows. Branston (1976) studied the variations of parameters in GQM model, using the *lognormal* followers headway distribution. Both authors have found evidence to support their assumption, that is, the followers headway is independent of traffic flows. These are good examples that parameters can be better understood and they can be used as important information in either model fittings or understanding of some traffic phenomena, no matter whether they are dependent or independent of the changes of traffic flows.

With more parameters or complexities in models, the fittings often behave more flexibly, and they might be sensitive to initial values. These problems sometimes don't result in failure in the GoF test, and often multi-sets of parameters can be accepted by GoF test. But this does not mean that the tested data samples are necessarily drawn from these distributions. When this happens, there is no method available for researchers to test further in order to choose a better approach.

However, the above can be overcome if the fitted parameters are more understandable with more cleared physical significance, as some estimations may exceed their reasonable ranges (or physical boundaries). To achieve this, both Wasielewski's and Branston's work is very valuable, as in their opinion, some parameters can be pre-determined before the actual fitting stage of mixed models. Potentially, when some of these parameters are pre-determined, models will not only be simplified with fewer fitting parameters, they will also be restricted. These may reduce the problems mentioned above. Moreover, with some parameters pre-determined by traffic conditions rather than by data samples, there are more reasons to believe that a model is a better presentation of headways.

The fourth problem is that, even with a good fitted model, it is still difficult to say whether the model can closely represent the headway process, especially the drivers behaviours of car following, which is one important purpose of headway analysis. So far, only the *gamma* and *lognormal* distributions are reported with good fitting quality in studies of mixed models, however, the lognormal-GQM is known to have very high computational cost[1] in fitting. The results in the *gamma* distribution as probably the only applicable followers headway model. Since no other good mixed models are available to compare with, researchers have to rely on the results of the gamma-mixed models, with the uncertainties of whether the models are close enough in representing the headway process.

[1]See footnote in chapter 2, page 48.

Overall, finding good statistical methods in headway analysis has still a long way to go, and all these issues need to be solved or bypassed if headway processes are needed to be better understood.

1.5 Objectives and Contributions

The first objective of this study is to propose and develop a software tool that will help on statistical headway analysis by providing common methods and data structures. A few functionalities shown as follows are the focus of this tool:

1. converting and managing headway data samples in common data structures;
2. providing mechanics to quickly develop and implement distribution models;
3. providing methods to perform goodness-of-fit tests in manner suitable for meaningful comparison;
4. storing and reporting data and fitting results for later analysis and comparison.

The tool is essential for efficiently appraising models and reporting results in a similar format. This makes the comparisons between models or datasets more intuitive.

The second objective is to study the parameters of mixed models, following similar methods suggested by Wasielewski and Branston, to investigate whether these methods will help in effectively simplifying and restricting the parameter estimations in fitting stages of mixed models. To fulfil this objective, several mixed models need to be discussed and implemented for comparison with findings reported by others.

Some analogies exist between car following and tasks used to measure human response times (RT) in cognitive psychology. Both have stages of reaction, decision making and action taking. Thus, some RT models might be suitable to apply in headway analysis. This may lead to alternative models to describe headway data. The third objective is to develop and implement some well-known RT models for the use of headway analysis, and to examine their fitting performance.

As a result of this work the following main outcomes and contributions to knowledge have been achieved:

1. A framework that helps to carry on the work of time-headway analysis is proposed in this study. This framework, which is developed using the MATLAB Statistics Toolbox, has several advantages in headway analysis. Firstly it assists in the process of data preparation and sampling by defining common data structures. Secondly, it quickly implements existing or newly developed models, and provides an interface to assist with parameter estimation and GoF tests. Thirdly, it manages all models, samples and results of analysis in one repository for further data visualisation and analytical report generation. The proposed framework is used for all headway studies reported within thesis. The study demonstrates that such a framework reduces the time and cost of repeatedly testing existing time-headway models and encourages researchers to focus on the important tasks of

model selection or development.

2. The gamma- and lognormal-mixed models of both SPM and GQM are tested using data collected on the M42 motorway near Manchester in the UK. The results show very similar fitting performance for these models compared with previous studies. In particular, the gamma-SPM model is further developed here, by deriving an explicit form of its CDF function, and this greatly improved the speed of GoF test of the model. Also a lognormal-SPM is for the first time implemented using approximation of the Laplace transform of the *lognormal* distribution. Furthermore, the use of pre-determined parameters was investigated. This study found that fitting performance can be greatly improved by using the suggested combinations of parameters, which are the vehicle arrival rates (λ) and selected parameters of the corresponding followers headway distribution, without great variations in the results of parameter estimations.

3. Two of well-known response-time distribution models, i.e. the exponential modified Gaussian (EMG) and inverse Gaussian (IVG) distributions, are discussed and tested as single models. They are found to provide acceptable fitting performance for the first two lanes of a three-lane motorway. The results are comparable to the shifted-lognormal single model, which has the greatest credibility as a single model in headway analysis. These two models are hence applied to define the followers headway distributions of mixed models (both SPM and GQM). The fitting results show equally excellent performance when using these new mixed models to fit the empirical headway data. This provides researchers with additional alternative models with good fitting performance in headway analysis.

Again, pre-determining some suggested combinations of parameters will simplify the processes of parameter estimation.

Two publications are currently in preparation based on the outcomes of this work, as outlined in Chapter 6, Section 3.

1.6 The Structure of This Thesis

Chapter 2 provides a detailed description of a proposed analytical framework that is developed in MATLAB (2013b) environment. This includes the definition of the structure of the framework, data structures, graphic user interface design and the use of parameter estimation and GoF methods. Chapter 2 also explains the approach of using *Statistics Toolbox* to implement headway models with both built-in and custom distributions. A set of tests on selected models are implemented, and comparison of results are presented and discussed.

Chapter 3 describes the headway dataset used in this study. Data sampling methods are explained, and basic statistics of processed headway samples are given.

Chapter 4 focuses on the traditional mixed models, for both SPM and GQM. These models are explained and tested, and results of analysis and comparison are shown. The focus then shifts to a very important feature of these two models, i.e. the Poisson λ. Based on this, the modified models are proposed with a pre-defined λ estimated from traffic flow rates. Then the fitting and GoF test results are shown. Further parameter simplification attempts are made and tested.

Chapter 5 presents the proposals of using response time (RT) models in headway analysis. Both selected RT models, i.e. the IVG (inverse Gaussian) and EMG (exponential modified Gaussian) distribution models, are explained. These two models are then added into the framework. The results of tests are presented and discussed. RT models are used as followers headway distributions for mixed SPM and GQM models. Methods used for parameter simplifications are also attempted and discussed.

The thesis concludes with a discussion of its findings, along with some plans for future work, in chapter 6.

Each chapter has a summary of contents at the beginning and a discussion at the end.

Chapter 2

A Framework for Time-Headway Statistical Analysis

Summary

This chapter proposes a framework that helps to simplify the procedures for testing new or existing headway distribution models. The advantages of using the MATLAB Statistics Toolbox to develop this framework are introduced, and the structure and data management of the framework are explained. As examples of using this framework, models are implemented and tested using empirical headway data. Finally, the test results are compared and discussed.

2.1 A Framework for Headway Data Analysis

Traditional ways of implementing a statistical model for headway analysis involve a substantial amount of work, which can be very tedious and time consuming. Such development procedures normally involve, more or less, the following summarized steps:

1. Developing necessary statistical functions, such as, probability density function (PDF), cumulative distribution function (CDF) etc.
2. Implementing parameter estimation methods, such as maximum likelihood methods, and goodness-of-fit test methods, such as, KS test, AD test, etc.
3. Preparing data with appropriate structures required by the above procedures.
4. Cleaning and sampling headway data.
5. Fitting and testing models against sampled data and collecting test results
6. Generating tables and graphs for presentation and analysis.

Even though one can design a systematic structure to simplify the repetition procedure of steps 3-6, however, the large amount of work involved in the first two steps for development of each new model is inevitable. This can be a barrier to many who wish to test rapidly some simple ideas. This also can be a barrier to some experienced researchers, as each has to repeat the same kind of work that has already been accomplished by others. These repetitions also involve variations in implementation due to the lack of standard procedures for model fitting and data sampling. These problems with non-standard procedures may happen in every step of development, and this makes comparisons between studies more difficult.

Moreover, there might be errors that are difficult to identify introduced during development, which may cause misleading conclusions.

To minimise the above problems, a concept of having a framework for headway analysis, which can process data and models more systematically and structurally, seems to have advantages. It can provide an opportunity to contribute new work on the same platform, to avoid unnecessary repetition, and hence to focus more on the actual model. It also provides a standard way of preparing data samples and dealing with models. The suggested framework in this chapter may not be the most convenient one, but it certainly provides a better and easier option until more comprehensive professional solutions are available. This work is still under development, and it is planned to be released for public access once completed.

2.2 Introduction of the Framework

2.2.1 Advantages of using MATLAB and Statistics Toolbox

MATLAB is an advanced mathematical modelling and analysis tool with many well-known highlights. To the extent of headway data analysis, a number of features are directly beneficial, and these are the main reasons for choosing MATLAB as a platform for developing our framework. The key advantages lie two main aspects, i.e. the Statistics Toolbox and object oriented programming capability.

The Statistics Toolbox provides comprehensive functionalities for work of statistical analysis. A few main features are (upto version 2013b, at time of this study) are highlighted below:

1. More proprietary distribution models become available over time as built-in models in the Statistics Toolbox.
2. Implementing new user-written distribution models in MATLAB is easy and straightforward.
3. Dataset arrays, provides powerful features for data management.
4. Many goodness-of-fit test methods are available as built-in functions.
5. Comprehensive tools are provided for designing graphical user interfaces and publishing analytical reports.
6. Object oriented programming features provide inheritable structures that support easier implementation of statistical models, data fitting, testing and analysis.

MATLAB continuously adds well tested distribution objects into the Statistics Toolbox; upto version 2013b, 22 continuous parametric probability distribution models are available as built-in models, and these cover all those commonly used single models introduced in the previous chapter. All these models are fully developed and tested, with comprehensive features to assist fast, easy and reliable implementation. This provides opportunities for researchers to test headway models with minimum coding effort. All models are provided with the same class structure by using the advantages of object oriented programming. Hence, all distribution models have common properties and methods that users can use directly without concerns over lack of standardisation.

Each probability distribution model can use the following inherited methods in common:

- **pdf** - Probability density functions
- **cdf** - Cumulative distribution functions
- **inv** - Inverse cumulative distribution functions
- **stat** - Distribution statistics functions
- **fit** - Distribution fitting functions
- **like** - Negative log-likelihood functions
- **rnd** - Random number generators

These common features and inheritance of classes provide maximum simplicity as all models are implemented in an almost identical way. Hence, the Statistics Toolbox provides opportunities for users to customize their models with exactly the same class structure. This customization can be simply achieved by inheriting from common parent classes identical to those built-in models provided in MATLAB. Such capability of customization is particularly useful for headway analysis, as many of the combined- or mixed-models have the same function structures that can be simply inherited from just one parent class. This substantially reduces the development work and makes newly developed models more robust and reliable. The mixed models discussed in chapters 4 and 5 are implemented using these beneficial features.

A comprehensive data type, i.e. *dataset arrays*, was introduced since the release of version of MATLAB R2007a, and this data type is particularly useful for managing and storing statistical data as objects. It provides a natural way to encapsulate heterogeneous statistical data and metadata, so that it can be accessed and manipulated using methods familiar to programmers that are analogous to those for numerical matrices. Data objects of this type can be simply stored, recalled or visually presented, and they are capable of managing all different data types, such as strings, arrays or even data objects, in a single data structure.

The Statistics Toolbox also provides over a dozen hypothesis test methods, and more will be available in due course. These cover the commonly used goodness-of-fit test methods in headway studies, such as Kolmogorov-Smirnov test (KS test), Anderson-Darling test (AD test) and Chi-square test (Chi-Sq. test). These methods are well compatible with all kinds of continuous distribution models; hence procedures of testing models become simpler and straightforward.

In addition to the above advantages, MATLAB provides powerful tools (GUIDE) for graphical interface design as well as tools (the Report Generator) that can generate and publish analytical reports.

The downsides of using MATLAB are mostly concerned with the following two aspects:

1. The cost of licensing is high comparing to other statistical analysis software.
2. Many new and useful features are not available in earlier versions of MATLAB.

2.3 Designs of the Framework

2.3.1 Structure of the Framework

Utilising the MATLAB features mentioned above, it is feasible to develop a framework that possesses three main aspects:

1. Capability of generalization to suit headway data structures obtained from various resources.
2. Quick and easy implementation of new customized models.
3. Well-defined methods for generating analytical reports.

To accomplish the above goals, the framework presented in Figure 2.1 is proposed.

The original headway data are normally collected using inductance detectors embedded under the surface of a road lane. These detectors record the vehicles' arrival time. Some of these devices have two detectors positioned nearby, hence they can also provide information of velocity and vehicle length. These data are fundamental for headway analysis. Other technologies, such as video cameras, are

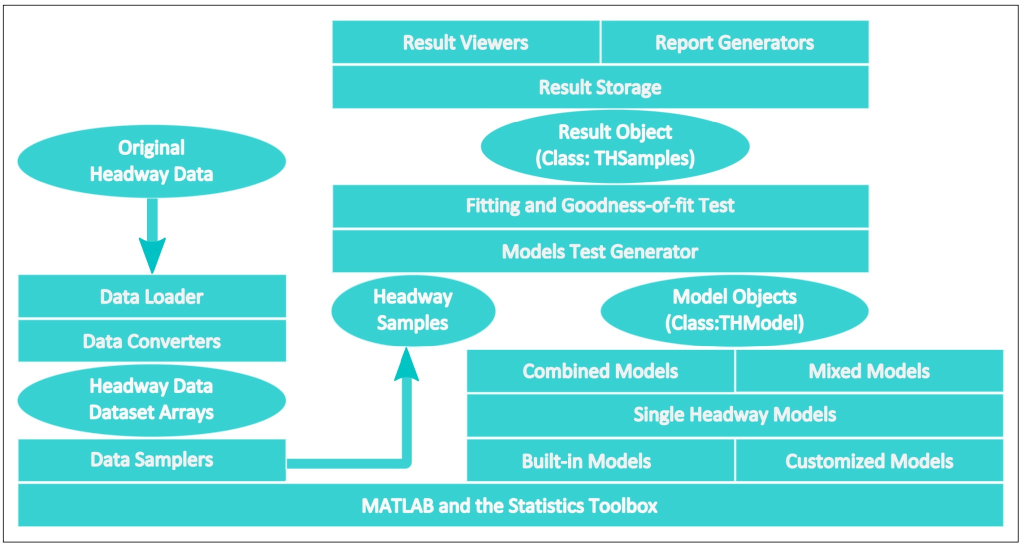

FIGURE 2.1: Structure of the proposed framework. After data conversion, further filtering and sampling will process data into required groups of samples that are then ready for model fittings. After data conversion, further filtering and sampling will process data into required groups of samples for model fittings

available for collecting similar headway data. Data collected using different technologies may have different data formats, and hence conversion from the original format is required to process them into common data structures before they can be used in this framework. The left part of diagram includes data loading and conversions that enable data to be prepared into the required format, which is specified in the next section.

After data conversion, further filtering and sampling will render data into required groups of samples that are then ready for model fitting.

All selected distribution models are managed as probability distribution objects, which have common class properties and methods. Hence, their objects can be operated in the same manner without special considerations given to any particular model.

The customized distribution models are those not included in the Statistics Toolbox, but they can be developed into the same structure as any built-in model. Both built-in and customized models can be used as base functions (e.g. as followers headway distributions) to formulate new kind of distribution. For instance, by combining location parameter δ a set of new distribution models can be derived as single or shifted-single headway models. In chapter 4, examples of using built-in models to form mixed headway models are explained in more detail.

The model objects and headway samples can be paired dynamically and stored as objects instantiated from a class so called `THTestprofile` class, and these objects can then comply parameter estimations and hypothesis tests. Result of such implementations together with these objects are finally stored as result objects that are instances of `THSamples` class.

The result objects are stored as files (.mat files), where they can be easily restored for further operation, such as viewing estimated parameters or GoF test results of any model, or generating analytical reports. These objects can also be modified to append or delete test objects by operating `THTestprofile` objects.

Using the above structure can simplify the procedures of data analysis, as data and distribution models are managed more systematically. Any changes in part of the data, adding new models or GoF test methods would not affect those tests previously accomplished. These changes can be managed in result objects where the new tests will be an extended part.

Researchers only need to concentrate on adding new distribution models, or test new data without concern over repeating work that has been pursued many times by others. The resulting objects and report generation capability make the work more readily comparable between studies, because they have been processed under the same structure and thus may thus be published in a similar format.

2.3.2 Specification of the Required Data Structures

The original collected headway data would normally have four important data fields, which are:

1. Arrival time of each individual vehicle
2. The lane of that vehicle
3. Vehicle Speed
4. Vehicle Length

Some of the data may have more information such as weather at the time of data collection, etc. These data are required to be formatted into proper data records of headway data that can be categorized for later data sampling and analysis. Hence, it requires a conversion of the original dataset to the appropriate structures.

A 15-field data structure is proposed in Table 2.1, which will help with further grouping and sampling the data. This data structure is intermediate but necessary for making consecutive samples of each lane of roads later, as well as keeping

TABLE 2.1: Proposed Structure of Headway Data

Fidel No.	Field Name	Description
1-4	*T, LID, V, Length*[1]	The original fields of data, arrival time, lane, Speed(m/s), and vehicle length(m)
5-8	*Day, Hour, Min, Sec*	Converted Time as Day, hour, Minutes and Seconds
9-10	*preTH, folTH*	Headway, and headway of followers (s)
10-11	*pdelV, fdelV*	Speed difference with front vehicle or following vehicle (m/s)
12-13	*preL, folL*	length(m) of front vehicle or following vehicle
15	*idx*	idx number in the current lane of road

original information such as speed difference which otherwise will be lost after sampling process.

Data conversion will be performed as the first step after data loading. For any original data format that has not previously been used in the framework, new data conversion functions may be required, as the structure of the original data may not be recognized.

The next step is data sampling, which will process and filter data using different techniques in order to achieve trendless and stationary headway data. More details about data sampling methods are discussed in chapter 3. An example of processed data can be found in Appendix A, Table A.1.

The data structure after the sampling stage is shown in Table 2.2, and this includes fundamental as well as statistics information that will assist categorization and preview of headway data samples.

The processed data with these two data structures are stored as dataset arrays in *'.mat'* file, which is a commonly used MATLAB file format for data storage.

Table 2.2: Proposed Structure of Data Samples

Field Name	Description	Field Name	Description
SampleID	Sample identification	Day	Day of sample
Size	Sample size (number of headways)	startHour	Hour of day
rateFlow	Flow level (veh/hour)	endHour	Hour of day
stdFlow	Standard deviation of Flows	Mean	Mean of headways
meanSpeed	Average Speed (mph)	Std	Standard deviation of headways
stdSpeed	Standard deviation of Speed	CV	Cofficents of variation
startTime	Starting time of sample	Skewness	Skewness of sampled headways
endTime	Ending time of sample	Kurtosis	kurtosis of sampled headways
LaneID	Lane identification	pLGV	Propotion of large goods vehicle
Sampleobj	data object containing original data		

2.3.3 The Interface Design

Graphical user interface (GUI) design includes modules that helps users to convert and manage headway datasets, to fit models, to perform goodness-of-fit tests and to report results. Figures 2.2 to 2.5 show examples of graphical interfaces that have been developed.

The main application window includes management interface of headway analysis projects, and buttons that call other modules.

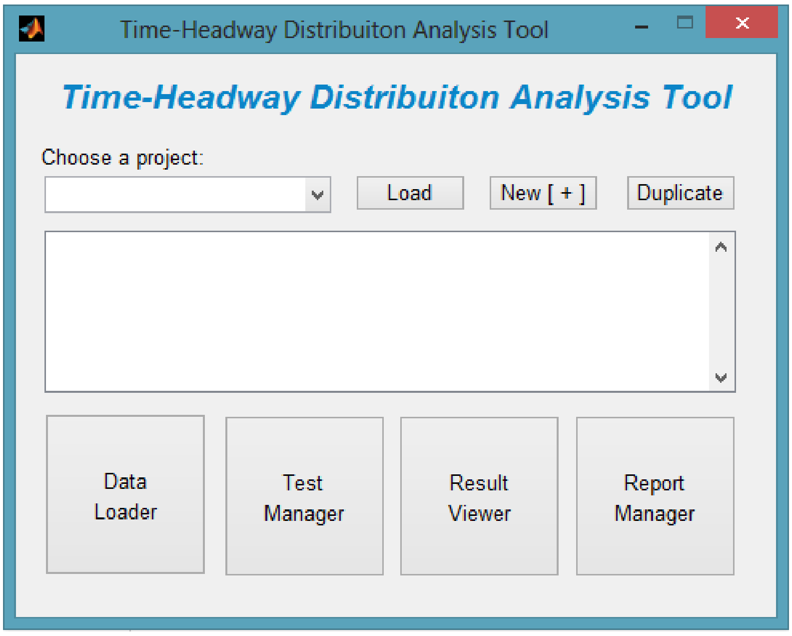

FIGURE 2.2: Main Application Window

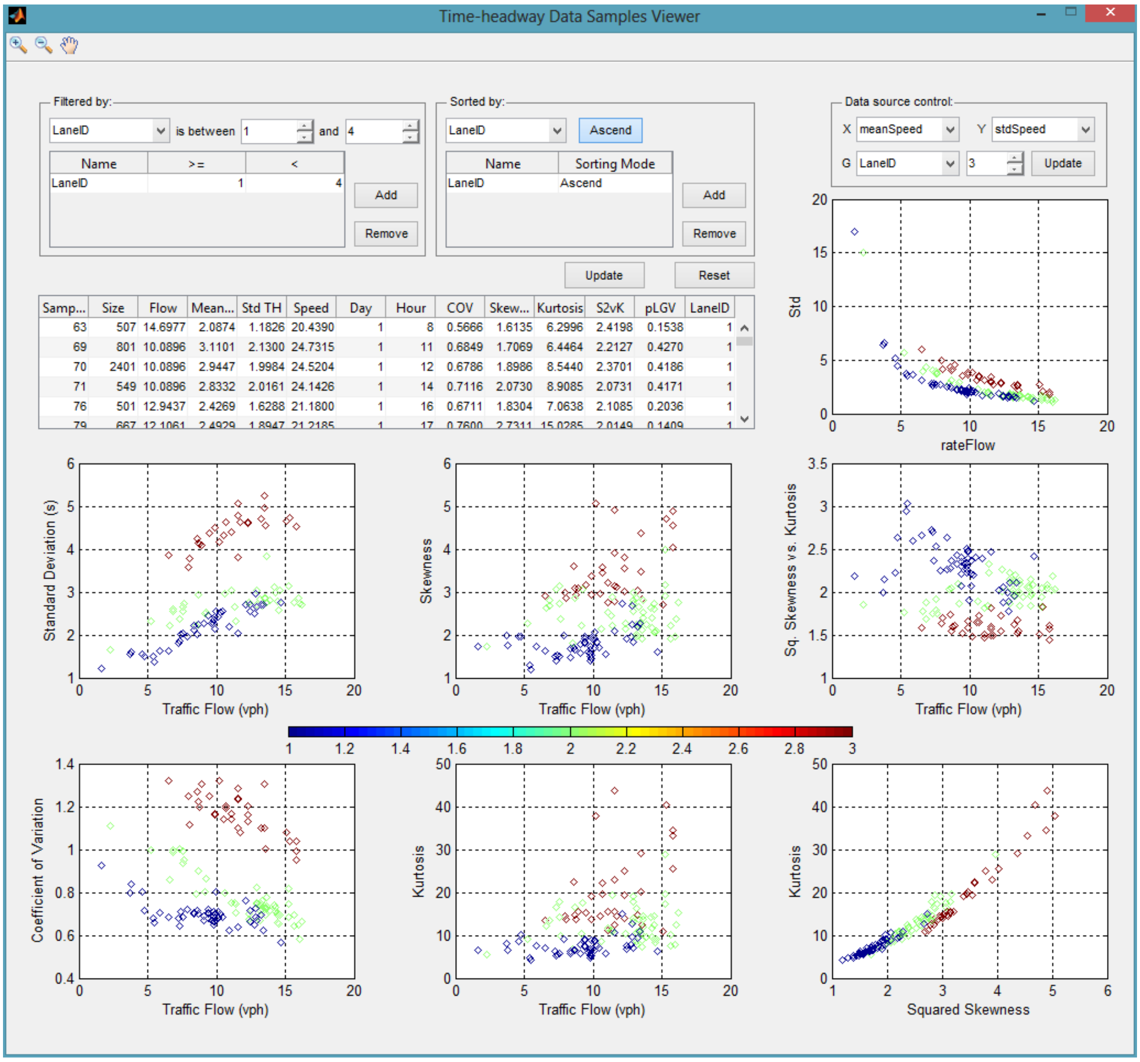

FIGURE 2.3: Samples Data Viewer Window

Samples viewer window presents data prepared in sampling process, and shows general statistics that summarize a group of chosen headway samples.

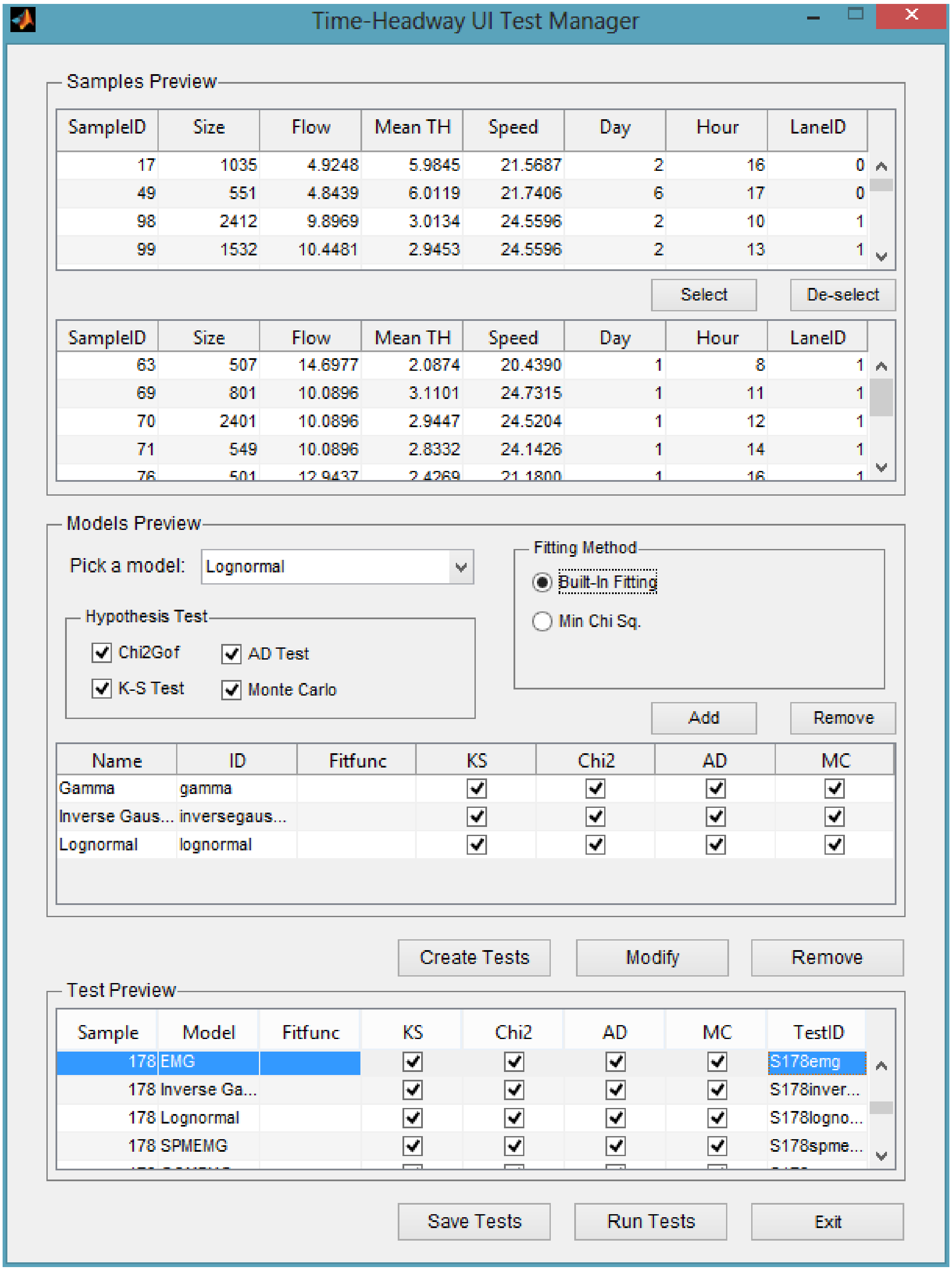

FIGURE 2.4: Headway Data Test Manger Window

The headway samples then can be processed selectively to test individual or a group of distribution models, as shown in Figure 2.3. In this step, the chosen models and samples will be paired and stored as tests that will be processed by using the button of 'Run Tests'. This will perform parameter estimation and goodness-of-fit tests for each pair of chosen data sample and distribution model. Figure 2.4 shows the uses of *gamma, lognormal* and *inverse Gaussian*(IVG) distributions as examples of tests.

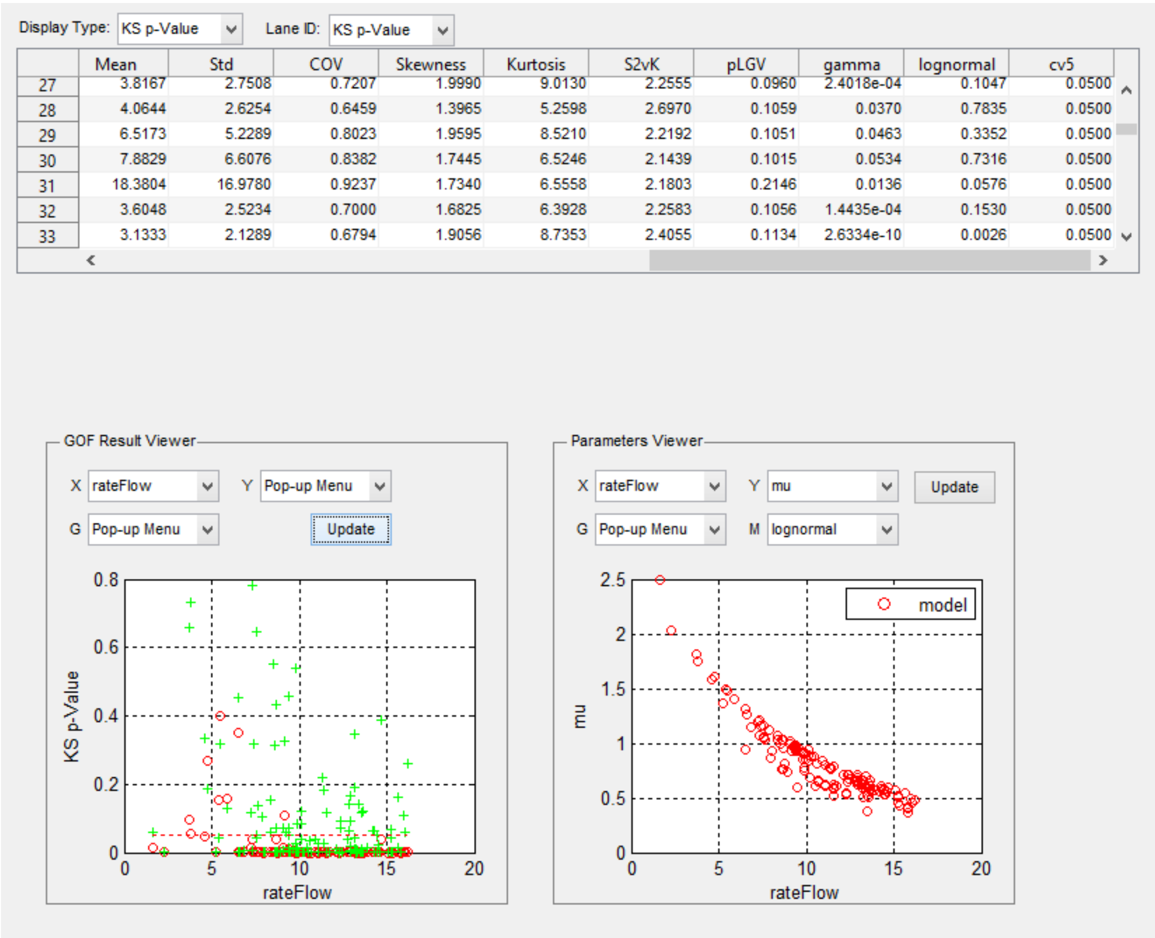

FIGURE 2.5: Group Results Viewer Window

The processed data are then stored in an object in the form of dataset arrays for further analysis. Figure 2.5 shows an example of data viewers, which are still under development, and such data viewers can help users to quickly review the result of individual or groups of data samples and models. Then all the results including statistics of data samples can be processed as tables and charts that will be stored in analytical reports for later reviewing or publishing.

2.3.4 Method of Adding a Customized Model

The Statistics Toolbox provides a sample of implementation of the *Laplace* distribution, and this example is summarized here to explain the method of creating a new model in MATLAB environment. A source code file LaplaceDistribution.m has been provided by MATLAB, and this can be used as a template to implement any customized distribution model. To make customized models recognizable by Statistics Toolbox, such model files have to be stored in a folder specially named as '+prob'. MATLAB will treat files contained in such folders as distribution models.

As an example, `LaplaceDistribution` class is a subclass of `prob.ToolboxFittableParametricDistribution`, which is parent class of all parametric distribution classes, in MATLAB. An object of the `LaplaceDistribution` class represents a *Laplace* probability distribution with a specific location parameter `MU` and scale parameter `SIGMA`. This distribution object can be created directly using the `MAKEDIST` function or fit to data using the `FITDIST` function. Any distribution object with these features may be implemented in the proposed framework.

Once the `LaplaceDistribution` class is created, some properties and methods need to be specified by overriding those inherited from its parent class. These methods and properties are listed in Tables 2.3 and 2.4.

Table 2.3: Inherited Methods For a Customized Model

Methods	Description
`cdf`	Cumulative distribution function
`fit`	Fit distribution to data
`likefunc`	Likelihood function
`icdf`	Inverse cumulative distribution function
`mean`	Mean
`median`	Median
`pdf`	Probability density function
`random`	Random number generation
`std`	Standard deviation
`var`	Variance

Table 2.4: Inherited Methods For a Customized Model, Using Laplace distribution as an example, with parameters mu and sigma specifically

Properties	Description
`DistributionName`	Name of the distribution
`mu`	Value of the mu parameter
`sigma`	Value of the sigma parameter
`NumParameters`	Number of parameters
`ParameterNames`	Names of parameters
`ParameterDescription`	Descriptions of parameters
`ParameterValues`	Vector of values of parameters

Note that the above methods and properties are required to create an optimum and fully functional distribution model, which may not be necessary when they are used in the framework for headway analysis. For instance, the `pdf, ...` `cdf, mean, std` and `var` are useful and they are required to be overridden, but methods like `icdf, random` or `median` can be ignored if they are not needed in headway analysis in some stages. The properties listed in Table 2.4 are only those important to this study. There are more properties in completing a distribution class, however, they are less relevant here. Properties `mu` and `sigma` are specially used for the *Laplace* distribution, and they should be changed correspondingly when using other models, such as `alpha` and `beta` for a *gamma* distribution.

Method `fit` and `likefunc` are important particularly for parameter estimation, and these two are frequently used in the framework. Further details will be introduced in the following section.

After the required methods and properties are coded properly, the customized model is ready as a functional distribution for headway analysis. They can then be fitted to headway data, and tested by GoF methods for example, which are explained in the following sections.

2.3.5 Parameter Estimation Methods

Parameter estimations are essential steps for calculating parameter values of a headway model, which are also specified as fitting models to data samples. To choose a good estimation method, fitting performance is an essential criteria, which include two main aspects, i.e. quality of fitting and estimation time. Fitting qualities describe how well the fitted models represent the sampled data, and they are normally measured by goodness-of-fit test, which is discussed in the next section. The estimation time is a computational cost that indicates how rapidly the fitting precess can be implemented. The fitting performance is the main method used in this study to measure a selected headway model.

There are a list of estimation methods used in fitting variety of headway models in the past. These can be known as moment estimator (ME) (Buckley, 1968; Cohen and Whitten, 1980), maximum likelihood estimator (MLE) (widely used by many), minimum of chi-squared statistics (MCS)(Branston, 1976), minimum of the sum of the squared errors (MSSE) between estimated and empirical distribution functions (Zhang et al., 2007) or minimum of the mean integrated squared error in the frequency domain (MISE) (Hoogendoorn and Botma, 1997). Or with modifications of some of above methods, such as, modified maximum likelihood estimator(MMLE)(Luttinen, 1996) or modification of ME (Ashton, 1971). Some studies also used a combination of certain of the above estimators for fitting different parts of a model (Branston, 1976). Luttinen (1996) and Ha et al. (2012) give detailed reviews and investigations of these methods.

The most frequently used estimation method is probably the traditional maximum likelihood estimator (MLE). Suppose θ ($\theta = \theta_1, ..., \theta_k, \theta \in \Theta$) are parameters of a parametric model with probability density function (PDF) of $f(t;\theta)$, and there is a sample $t_1, t_2, ..., t_n$ for n independent and identically distributed (iid) headway observations. Assuming the sample can be described by the distribution model $f(t;\theta_0)$, then θ_0 is referred to the *true value* of the parameter, which is unknown. The purpose of fitting a model is to find an estimator $\hat{\theta}$ which would be as close to θ_0 as possible. The method of MLE is supposed to maximize a joint density function of all observations of the given data sample, in which the joint density function can be expressed as:

$$f(t_1, t_2, ..., t_n; \theta) = f(t_1; \theta) \times f(t_2; \theta) \times \cdots \times f(t_n; \theta). \tag{2.1}$$

If considering the data sample $t_1, t_2, ..., t_n$ as fixed parameter, and θ as variables in the above function, then the likelihood function can be written:

$$\mathcal{L}(\theta; t_1, t_2, ..., t_n) = f(t_1, t_2, ..., t_n; \theta) = \prod_{i=1}^{n} f(t_i; \theta). \tag{2.2}$$

This function is more convenient to use as its logarithm, i.e. the log-likelihood:

$$\ln \mathcal{L}(\theta; t_1, t_2, ..., t_n) = \sum_{i=1}^{n} \ln f(t_i; \theta). \tag{2.3}$$

The MLE estimates $\hat{\theta}$ to the closest value of θ_0, the true value of the parameters, by maximizing the likelihood or log-likelihood function, i.e.:

$$\left\{\hat{\theta}\right\} \subseteq \left\{\underset{\theta \in \Theta}{\arg\max} \ln \mathcal{L}(\theta; t_1, t_2, ..., t_n)\right\}. \tag{2.4}$$

Often with implementation of MLE in MATLAB, the negative log-likelihood function is more convenient for optimization

$$\hat{\theta} \subseteq \underset{\theta \in \Theta}{\arg\min} \left\{- \ln \mathcal{L}(\theta; t_1, t_2, ..., t_n)\right\}. \tag{2.5}$$

The log-likelihood function sometimes can be resolved for $\hat{\theta}$, if solutions exist for its first order derivative of each parameter to equal zero:

$$\left.\begin{aligned} \frac{\partial \ln \mathcal{L}(\hat{\theta})}{\partial \hat{\theta}_1} &= 0 \\ &\vdots \\ \frac{\partial \ln \mathcal{L}(\hat{\theta})}{\partial \hat{\theta}_k} &= 0. \end{aligned}\right\} \tag{2.6}$$

However, these equations may not always be possible to solve with explicit solutions, as there are cases where distribution models have very complex form. In such cases, numerical methods have to be used. Lists of studies have discussed and recommended the MLE method for model fitting of headway data (Cohen and Whitten, 1980; Ha et al., 2012; Luttinen, 1996; Stuart and Ord, 1991).

Eq. 2.5 can also be directly used numerically to find $\hat{\theta}$, and this is a more generalized way compared with using partial derivatives in Eq. 2.6, as the first method does not depend on any specific form of $f(t)$. That is, as long as the PDF of a custom distribution is available, MLE can be implemented using Eq. 2.5 to fit sampled data. This numerical MLE method is commonly used in MATLAB and other statistics software packages, such as R. In Statistics Toolbox, function `mle` will use MLE to estimate $\hat{\theta}$ this way, when a gradient of a custom distribution is not available.

In the case of using the proposed framework, this numerical MLE method is extremely useful, as it can help fit any customized distribution model to headway data, without dealing with details of mathematical properties of those models.

Moreover, this method is not the only one used by `mle` function, which is also able to identify the best fitting method if available in the distribution object. For instance, `mle` function will not implement a numerical method of a *normal* distribution, as it has explicit solution of Eq. 2.6 that is available in the `prob.NormalDistribution` class. This useful feature is used in the proposed framework, and it provide flexible manner in fitting models.

The above are the main methods used for parameter estimation in the framework. In summary, for any models (mostly single models) which have been developed previously, e.g methods provided in MATLAB, their fitting methods will be used. However, with more complicated models, such as mixed distributions introduced in chapters 4 and 5, numerical MLE method is used to directly estimate parameters. This feature of the framework will minimize the work related to fitting headway models, especially when researchers would wish to test the fitting quality of a model, before diving into a great deal of model optimization. However, this does not limit its capability if any optimized fitting method is developed, as these methods can be coded to override the `fit` method of the custom distribution class, and it will then be used by the framework.

Note that the numerical MLE method computes the parameter estimates using an iterative maximization algorithm. In such case, with some models and data, poor choices of initial values can cause this method to converge to a local optimum rather than the global maximizer, or to fail to converge entirely. Hence the choice of initial values are often crucial, and this has to be treated with more care.

2.3.6 Goodness-of-fit (GoF) Tests

To test whether a model is suitable to describe sampled headway data, two methods are commonly used, i.e. using a visual examination or applying a goodness-of-fit test. The two types of method are often used together. A graphic of curves of probability density function (PDF) or cumulative distribution function (CDF) are presented together with histograms or empirical distribution function respectively for comparison. Some researchers also used Quantile-Quantile (Q-Q) plots for visual exam (Zhang et al., 2007). These graphical methods are subjective and they are only suggested to use for preliminary evaluations.

A goodness of fit test is a statistical test in which the validity of one hypothesis is tested without specification of an alternative. The general procedure consists in defining a test statistic, which is some function of the data measuring the distance between the hypothesis and the data (Bock and Krischer, 1998). These tests are used to determine whether the outcome of a study would lead to a rejection of the null hypothesis based on a pre-specified low probability threshold, which can to decide if a result contains sufficient information to cast doubt on the null hypothesis (Schlotzhauer, 2007). To be clear, following the conversion of MATLAB, a p-value of a test is defined to reflect the probability of observing a test statistic as or more extreme than the observed value under the null hypothesis. P-values are often coupled to a significance level, which is also set ahead of investigation, usually at 0.05 (5%). Thus, if a p-value was found to be less than 0.05, then the result would be considered statistically significant and the null hypothesis would be rejected (McKillup, 2005). Other significance levels, such as 0.1 or 0.01, are also used, depending on the field of study.

Particularly to this study, GoF tests are mainly used in testing the null hypothesis:

$$H_0 : \text{The random headway sample } \{t_1, ..., t_n\} \text{ is generated by } F(t|\theta) \tag{2.7}$$

where $F(t|\theta)$ can be any specific distribution of interest in this study. A significance level of 0.05, rather than 0.01, has been chosen to make the hypothesis testing more restrictive, since this study is to find a model that is more suitable to describe the headway data.

Three types of GoF methods had majorly attentions in headway studies: 1) chi-square test(NIST et al., 2001), 2) Kolmogorov-Smirnov (KS) test(Jr., 1951; NIST et al., 2001; Stephens, 1974), and 3) Anderson-Darling (AD) test(NIST et al., 2001; Stephens, 1974). Luttinen (1996) suggested many of these tests are problematic, and he proposed that more powerful GoF test methods should be used, such as combination of probabilities or moving probability.

In the present study, the two GoF tests described below are used mainly to illustrate the fitting performance of models.

Kolmogorov-Smirnov (KS) test

The KS test is a nonparametric test of the null hypothesis that the population CDF of the data is similar enough to the hypothesized CDF. The test statistic D is the maximum absolute difference between the empirical CDF (also called EDF) calculated from t and the hypothesized CDF (the proposed model):

$$D = \max_{t \in Re_+} \left(\left| \hat{F}(t) - F(t|\theta) \right| \right) \tag{2.8}$$

where $\hat{F}(t)$ is the EDF for a sample $\{t_1, ..., t_n\}$, calculated from t:

$$\hat{F}(t) = \left\{ \frac{i}{n} \middle| t_{(i)} \leq t < t_{(i+1)} \right\}, \quad i = 1, ..., n. \tag{2.9}$$

Hence, the KS test statistics can be defined as (NIST et al., 2001):

$$D = \max_{1 \leq i \leq n} \left\{ F(t|\theta) - \frac{i-1}{n}, \frac{i}{n} - F(t|\theta) \right\}. \tag{2.10}$$

The p-value of KS test can bu calculated using:

$$p = 1 - F_{KS}(D|n), \tag{2.11}$$

where F_{KS} is the CDF of the KS distribution.

The KS test is nonparametric test as its test statistic is the independent of all hypothesized distributions and their parameters.

Anderson-Darling (AD) test

The Anderson-Darling test (Stephens, 1974) is used to test if a sample of data came from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than does the

K-S test.The test statistic belongs to the family of quadratic empirical distribution function statistics, which measure the distance between the hypothesized distribution, $F(t|\theta)$ and the EDF, $\hat{F}(t)$ as

$$Q = n \int_0^\infty \left(\hat{F}(t) - F(t) \right)^2 w(t) \mathrm{d}F(t), \tag{2.12}$$

over the ordered sample values $t_1 < t_2 < ... < t_n$, where $w(t)$ is a weight function and n is the number of data points in the sample.

The weight function for the Anderson-Darling test is

$$w(t) = \left[F(t)(1 - F(t)) \right]^{-1}, \tag{2.13}$$

which places greater weight on the observations in the tails of the distribution. The Anderson-Darling (AD) test statistic (A^2) is

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln(F(t_i)|\theta) + \ln(1 - F(t_{n-i+1}|\theta)) \right], \tag{2.14}$$

where $t_1 < t_2 < ... < t_n$ are the ordered sample data points and n is the number of data points in the sample.

The framework provides both visual examination and GoF tests. For visual presentation, PDF curves are often compared with histograms to show general impressions of fitting result. GoF tests are used to give more quantitative measures (p-values) of quality of fitting for models. MATLAB Statistics Toolbox offers a selection of GoF test methods, and for the purpose of headway analysis, the chi-square test, KS test and AD test are selected which can be used in the framework. These methods are provided as function of `chi2gof, kstest` and `adtest` correspondingly. The use of these tests are totally independent to the type of distribution models selected, and they are selected and operated by `THTestprofile` objects. Hence, it is easy to add new GoF test methods into the framework, by adding the function name into the list of GoF methods in `THTestprofile` class. For example, as shown in Figure 2.4, a 'Monte Carlo' GoF test method is added there, which is using a parametric KS test method suggested by Luttinen (1996).

2.3.7 Development Progress of the Framework

The framework is not yet fully complete, although data processing, sampling and model management element are fully functioned. Parts of the graphical interface on result viewer and report generators are still under development. The completed part of the framework can perform headway analysis work and it has been used to carry out the data processing and analysis work of this study. The next section gives an example of the use of this framework to test two selected built-in models, and presents some of the headway analysis results.

2.4 Examples of Implementation of Chosen Models

Implementations of any distribution in the proposed framework are straightforward, since it will automatically identify all models that have been developed and stored in folders with special name `+prob` mentioned earlier. Figure 2.4 shows controls that will help on choosing models, GoF methods and data samples. Note that the fitting method box is still under development, and the framework will automatically use the fitting method described in the overridden `fit` method of each model. In future, it is planned to allow more fitting methods to be chosen from a list, and this will help to enable a comparison to be made.

All the data preparation work is processed using the framework discussed in the above sections. Details of data descriptions and preparations are given in chapter 3.

2.4.1 Selected Probability Distribution Models

Perhaps the two most commonly studied single models are *gamma* and *lognormal* distributions, and many researchers had investigated them. The *gamma* distribution is popular due to its flexibilities (Ha et al., 2012; Luttinen, 1996; Zhang et al., 2007) related to other models, such as *exponential, Pearson III, Erlang* distributions, while the *lognormal* distribution are reported to have close connections to car following models (Greenberg, 1966). These two models, especially when they are modified with location parameter δ, have relatively better fitting performance to headway data as single models. Moreover, the mixed models are often studied with these two distributions as followers headway models. These mixed models are reported to give excellent fitting performance.

For the above reasons, *gamma* and *lognormal* models are selected as examples of implementation under the proposed framework. This demonstrates the capability of using such frameworks, as claimed in previous sections, and also enables the performance of both models to be compared with results from other studies.

In following section, *gamma* and *lognormal* distribution models, together with their shifted modifications are reviewed and tested.

2.4.2 The Gamma and Lognormal Distributions

Gamma and Shifted-Gamma Distributions

The rationale of using *gamma* distribution for describing headway distributions, is discussed in detail by Luttinen (1996); May (1990). Some mathematical properties relevant to this study are summarized below.

To clarify notations, PDF functions are notated with function form $g(t)$ for single models, while CDF functions are notated with $G(t)$. The headway random variable is commonly denoted X. These notations are applied across the entire thesis, and are used to distinguish functions of mixed models.

The probability density function (PDF) of *gamma* distribution is:

$$g(t) = g(t; \alpha, \beta) = \frac{\beta^{\alpha} t^{a-1}}{\Gamma(\alpha)} e^{-\beta t}, t > 0 \tag{2.15}$$

where $\alpha, \beta \geq 0$, and $\Gamma(a)$ is the gamma function

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} \mathrm{d}x \tag{2.16}$$

The cumulative distribution function (CDF) of *gamma* distribution is:

$$G(t) = G(t; \alpha, \beta) = \frac{\gamma[\alpha, \beta t]}{\Gamma(\alpha)}, t > 0 \tag{2.17}$$

where $\alpha, \beta \geq 0$, $\gamma(\alpha, x)$ is the incomplete gamma function

$$\gamma(a, b) = \int_0^{b} x^{a-1} e^{-x} \mathrm{d}x. \tag{2.18}$$

The mean and variance of *gamma* distribution is:

$$E[X] = \frac{\alpha}{\beta} \tag{2.19}$$

$$\mathrm{Var}(X) = \frac{\alpha}{\beta^2} \tag{2.20}$$

The *gamma* distribution has Laplace transform of:

$$g^*(s) = \left(\frac{\beta}{s + \beta}\right)^{\alpha} \tag{2.21}$$

The Laplace transforms of models are interesting for this study, mainly due to their important role in the SPM mixed models. Also, the Laplace transform of the distribution model is similar to moment generating functions, which are useful in deriving equations of moments for various purposes.

Applying a location parameter δ, the PDF and CDF of *shifted-gamma* distribution can be expressed respectively, as:

$$g(t) = g(t; \alpha, \beta, \delta) = \frac{\beta^{\alpha}(t-\delta)^{a-1}}{\Gamma(\alpha)} e^{-\beta(t-\delta)}, t \geq \delta \tag{2.22}$$

$$G(t) = G(t; \alpha, \beta, \delta) = \frac{\gamma[\alpha, \beta(t-\delta)]}{\Gamma(\alpha)}, t \geq \delta \tag{2.23}$$

where $\gamma(a, b)$ is the incomplete gamma function, $\alpha, \beta \geq 0$.

The *shifted-gamma* distribution is also known as Pearson type III distribution.

2.4.2.1 Lognormal and Shifted-lognormal Distributions

Using *lognormal* distribution in headway analysis is firstly proposed by (Daou, 1964, 1966) and Greenberg (1966). Many studies covered the applications or reviews of applying *lognormal* model (Branston, 1976; Gerlough and Huber, 1975; Luttinen, 1996; May, 1965; Tolle, 1971). The following part summarizes key mathematical properties of importance in this study.

The PDF of *lognormal* distribution:

$$g(t) = \frac{1}{t\sqrt{2\pi}\sigma} e^{-\frac{(\ln t-\mu)^2}{2\sigma^2}} \tag{2.24}$$

The CDF of *lognormal* distribution:

$$G(t) = \frac{1}{2}\left[1 + \text{erfc}\left(\frac{\ln t - \mu}{\sqrt{2}\sigma}\right)\right] \tag{2.25}$$

where **erfc**$(t)$is the complementary error function, $\mu, \sigma > 0, t > 0$.

The mean and variance of *lognormal* distribution are:

$$E[X] = e^{\mu + \frac{1}{2}\sigma^2} \tag{2.26}$$

$$\text{Var}(X) = \left(e^{\sigma^2} - 1\right) e^{2\mu + \sigma^2} \tag{2.27}$$

A general analytic Laplace transform of the *lognormal* distribution does not exist.

Applying a location parameter δ, the PDF and CDF of *shifted-lognormal* distribution can be expressed:

PDF:

$$g(t) = \frac{1}{(t-\delta)\sqrt{2\pi}\sigma} e^{-\frac{(\ln(t-\delta)-\mu)^2}{2\sigma^2}} \tag{2.28}$$

CDF:

$$G(t) = \frac{1}{2}\left[1 + \text{erfc}\left(\frac{\ln(t-\delta) - \mu}{\sqrt{2}\sigma}\right)\right] \tag{2.29}$$

2.4.3 Parameter Estimation

The parameters of *gamma* and *shifted-gamma* distributions are often estimated using methods of ME, MLE, MMLE and MCS. The ME method is reported to give poor performance for *gamma* distribution (Bownan and Shenton, 1982; Cohen and Whitten, 1988; Greenwood and Durand, 1960). Cohen and Whitten (1988) suggested modified method of moments estimators (MMMEs), and they claimed that the method is straightforward and applicable over the entire parameter space. MCS is used by Ha et al. (2012), and he suggested a three-step optimization method using MATLAB. The MLE methods are widely used and favoured by most researchers. Johnson et al. (1994) observed unstable performance with the estimated α close to unity, and he recommends that the method should only be used if expected estimation of α is more than 2.5. This suits most of the conditions in headway analysis. Modified MLE (MMLE) is proposed and used by Bownan and Shenton (1982); Luttinen (1996), who found acceptable performance when α is between 0.25 and 2.

Cohen and Whitten (1980) compared ME, MLE and MMLE for *lognormal* distribution, and they found estimation using ME was not as accurate as MMLE. Luttinen (1996) suggested convergence problems with the MLE method, despite the fact that it gave most accurate results. Hence he recommended the MMLE method for *lognormal* distribution. Zhang et al. (2007) suggested the use of MLE method to formulate the best unbiased estimators for single models in his study.

For *gamma* and *lognormal* distributions, the fitting methods are provided in the Statistics Toolbox. In contrast, parameters of *shifted-gamma* and *shifted-lognormal* are estimated using the numerical MLE method introduced earlier in this chapter. As discussed, numerical MLE is not assumed to be the best method, but it is the easiest method for rapid deployment.

The comparisons among different methods of parameter estimation are out of the scope of this study, but it is certainly worth considering as this framework is developed in future.

As discussed earlier, KS and AD GoF tests are used to examine the fitting performance of headway models in this study.

2.4.4 Results and Discussion

Estimation times[1] are measured during the tests and are summarized in Table 2.5. To distinguish between using location parameter or not, they are mark as 'Shifted' and 'Original' respectively. The estimation time of both models seem quite robust, with the average values about 0.2 seconds. *Gamma* fittings are slightly slower than *lognormal* models. The fitting of *shifted-Gamma* and *shifted-lognormal* are both slower than original ones, and these are expected as the additional parameter δ is introduced to the original models.

TABLE 2.5: Estimated times (s) on estimation of gamma and lognormal distributions

	Gamma		Lognormal	
	Mean	Std	Mean	Std
Shifted	0.24	0.05	0.17	0.02
Original	0.18	0.03	0.14	0.02

Table 2.6 shows the results of GoF tests. The p-values of both KS and AD tests are listed in the table. The values in brackets represent number of samples that did

[1]An estimation time is a measure of a computational cost of a parameter estimation. This time is measured by recording the elapsed cpu-time, which is the difference between the recorded cpu-times before and after the parameter estimation, in seconds. The technical features of the PC: Intel(R) Core(TM)2 Quad CPU Q9550 @ 2.83GHz; 4.0GB RAM.

not pass the corresponding test. Figures 2.6 and 2.7 show more detailed results of p-values against levels of traffic flows. All results are categorized with three lanes (marked as lane 1-3).

TABLE 2.6: Summary of GoF Test Results for Gamma and Shifted-Gamma Distributions

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
Gamma						
Shifted	21(31)	21(31)	4(43)	4(43)	0(33)	0(33)
Original	7(45)	5(47)	1(46)	0(47)	0(33)	0(33)
Lognormal						
Shifted	48(4)	49(3)	44(1)	44(1)	1(31)	0(32)
Original	29(23)	29(23)	24(23)	23(24)	0(33)	0(33)

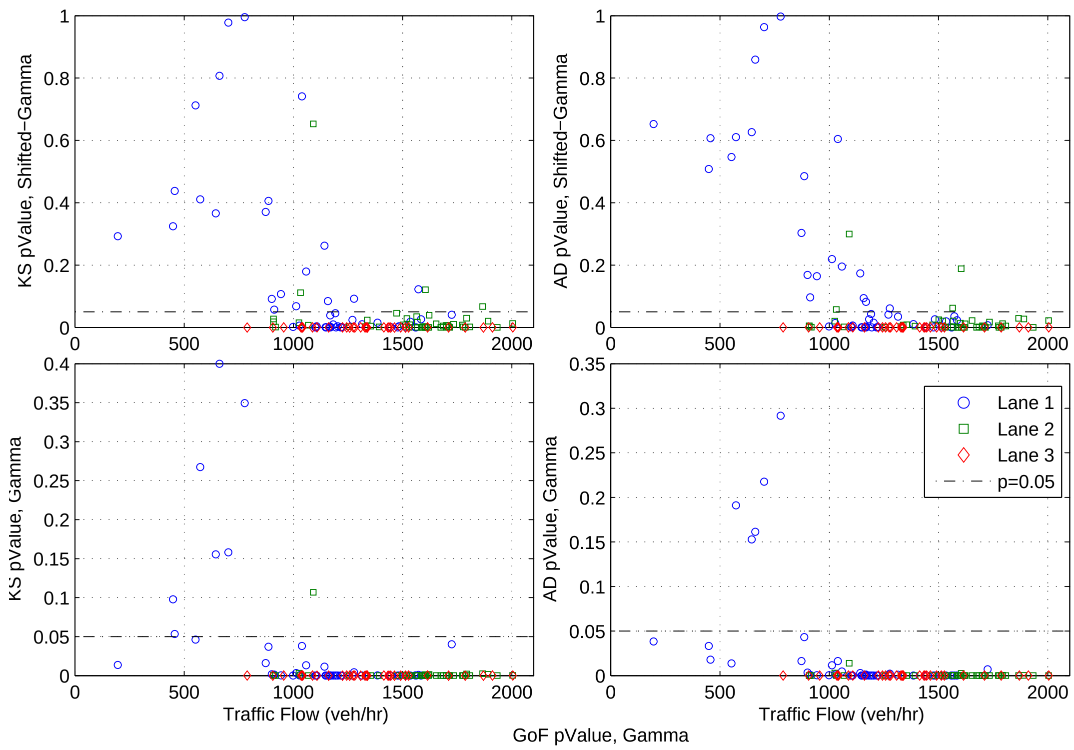

FIGURE 2.6: GoF Test Results On Gamma and Shifted-Gamma Distributions

The *gamma* distribution without location parameter δ has very poor fitting performance, as only 8 out of 132 samples have passed KS test, while only 5 have

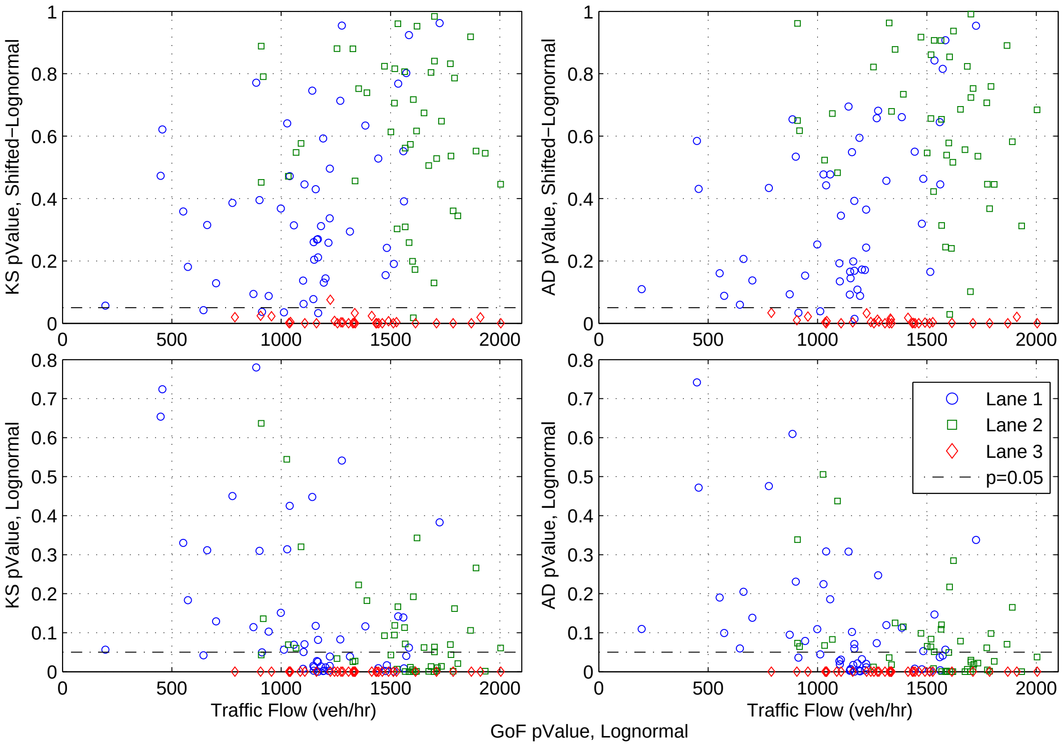

FIGURE 2.7: GoF Test Results On Lognormal and Shifted-Lognormal Distributions

passed AD test. The *shifted-gamma* has slight improvement, with 25 samples passed both tests. Figure 2.6 shows that most of the succeeded samples are from lane 1, at lower traffic flow levels (mostly under 1000 veh/hr). Similar poor fitting performance were reported by Luttinen (1996), in his thesis.

The *lognormal* distribution is commonly considered as the best single model reported by many previous studies. Result here shows indeed much better performance comparing to both *Gamma* and *shifted-Gamma* models. The best fitting performance are shown with *shifted-lognormal* models, as 92 samples have accepted by GoF tests, which is a big improvement compared with test result of original *lognormal* distribution. Most of the accepted samples are in lane 1 and 2, and better fitting performance is obtained with higher flows of traffic. However, it should be noticed that three samples failed in parameter estimation stage, as estimations are beyond the maximum number of iterations.

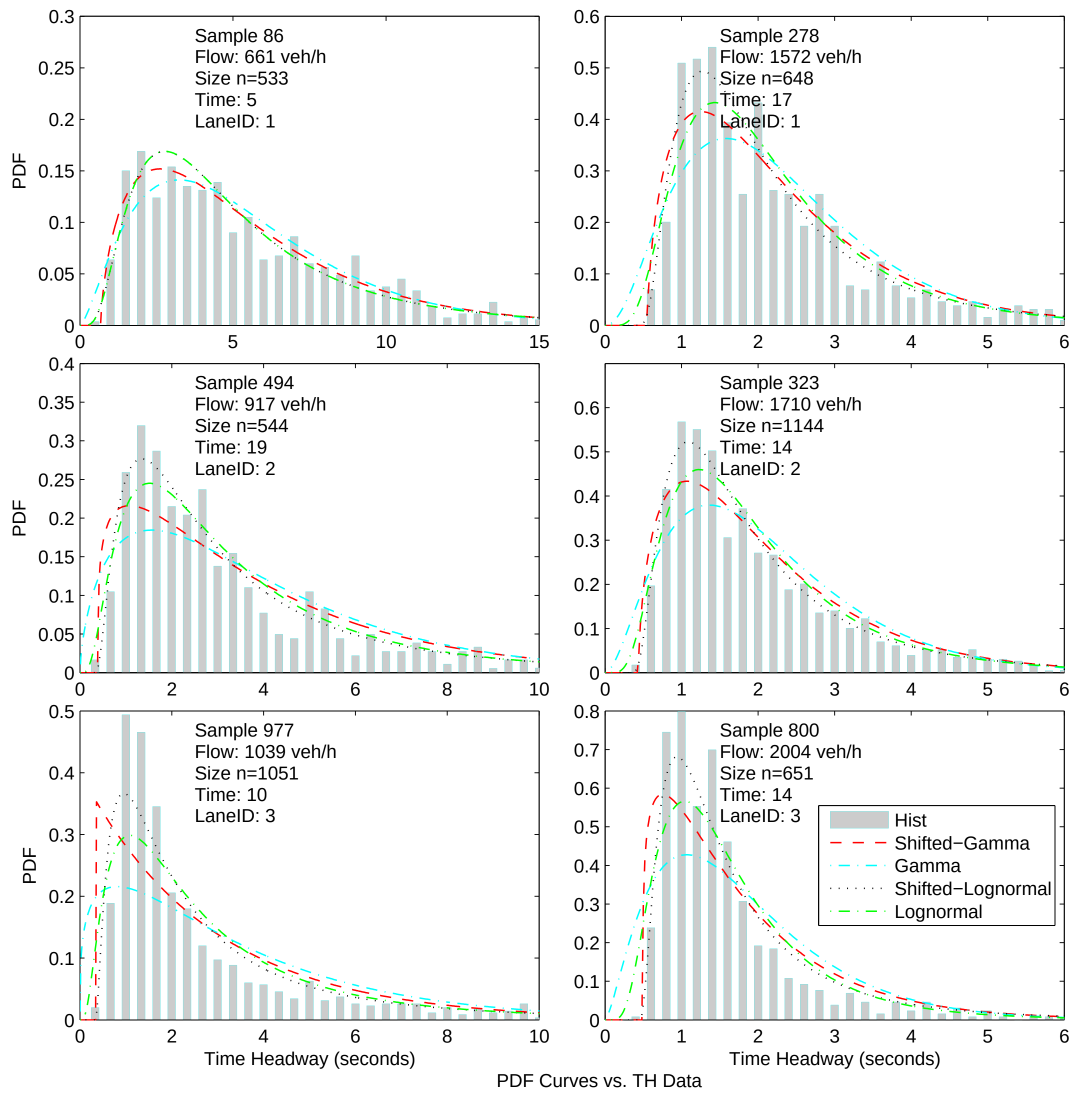

Figure 2.8: PDF fitting results of Gamma, shifted-Gamma, lognormal and shifted-lognormal on selected data samples

All the above models show very poor fitting results in lane 3, as only 1 sample using *shifted-lognormal* is managed to pass KS test. Referring to Figure 3.3, the lane 3, as the third and fast lane in motorway, has more complex traffic situations. The more skewed headway data indicate the existence of larger portion of short or very short headways comparing with first two lanes. As Ha et al. (2012) discussed, the single models, in general, have difficulty modelling short and very short headways.

Similar results can be seen later in chapter 5, when more single models are tested for fitting the third lane data.

Fitting results can also be viewed and compared visually. Selected results of PDF function compared with histogram of empirical headway data are shown in Figure 2.8. Six samples are selected in this figure, as two samples per road lane. The row 1-3 matches lane 1-3 respectively. The left column of figure contains samples with relative lower level of traffic flows, while the right column shows samples with higher traffic flows. Again, these figure show in general better fitting results with *shifted-lognormal* distribution. But for the lane 3, none of the models seem to fit the histogram well.

2.5 Chapter Discussion

A framework is introduced here which is intended to simplify the work on headway data analysis. Some of the advantages of using MATLAB as development software are discussed. The framework structure and data structure required are explained in detail to show how the proposed framework for headway analysis is implemented.

Two single headway distributions are implemented and tested using the proposed framework, and results have illustrated how the comparison of the selected models may be made. The results show that the *lognormal* model generally has better fitting performance compared with the *gamma* distribution model. The *shifted-lognormal* model shows the best fitting result.

Later chapters examine existing or new models that offer potentially good fitting performance to headway data. The data samples used for these tests are described in chapter 3. All data processing, implementation and tests is performed under the framework described here.

Chapter 3

Headway Data Preparation

Summary

This chapter describes the headway data used in this study. This is followed by an explanation of the data sampling process. Then the selected samples are introduced and their statistics are considered.

3.1 Data Description and Collection

Headway data used in this study were supplied, with acknowledgement, by Professor Eddie Wilson, Univeristy of Bristol, UK (Wilson, 2008), who studied the data to reform individual vehicle trajectories. The data are originally collected using the English Highways Agency's Active Traffic Management (ATM) system which operates on a 15-km (approx. 9-mile) stretch of the M42 motorway constituting part of the box of motorways around Birmingham. For the purpose of monitoring traffic closely, inductance detectors have been installed more densely, with a nominal spacing of 100 m, and more dense spacing of 30 m in some parts.

The data that each inductance detector records are listed in Table 3.1. The data are extracted using waveform analysis, with the arrivals quoted to 0.1 seconds accuracy.

Table 3.1: Data Fields and Nominal Resolution for Each Individual Vehicle Record (Wilson, 2008)

Quantity	Arrival Time	Lane	Velocity	Length
Resolution	0.1 s	integer 0-3	0.01 m/s	0.01 m

Only part of the data above is used in this study, and it is collected from a single location about 1 mile north of junction 4, on M42 northbound. It contains 4 lanes, which are defined as lane 0 to lane 4, this include the emergency lane (lane 0) that was partially used on at peak times. The lanes 1, 2 and 3 represent the slow, middle and fast lane of the road respectively. The data are collected in a continuous period from 1st to 7th October 2008. It is assumed that the data are free from effects caused by road ramp or other road configuration issues.

The dataset used here represents approximately 464,000 vehicles total in all four lanes, and Figure 3.1 shows an example flow level of lane 2 for the whole 7-day period.

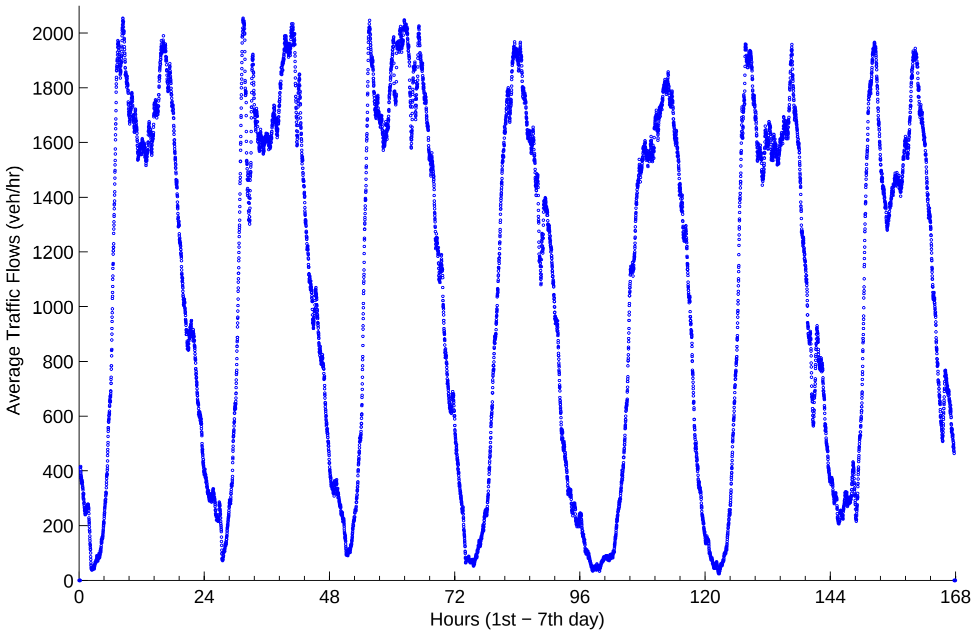

FIGURE 3.1: Average Traffic Flows of 7-day Period

3.2 Data Sampling Process

3.2.1 Methods of Headway Data Sampling

To carry out reliable analysis, the studied headway data should be under the condition of relatively stable traffic situations, which means both changes on velocity and flows are independent of time, so called stationary samples. Researchers proposed a few methods to test for more stationary samples.

Breiman and Lawrence (1973) discussed the approaches of achieving headway samples of stationary traffic flow. They attempted to separate fluctuations from the changes in flow rate and to estimate the parameters of the fluctuations. As a result, they proposed and recommended the method of area test.

Three trend tests methods are often used in data sampling process:

1. Weighted sign test (Cox, D. R. and Stuart, 1955)
2. Kendall's rank correlation test (Kendall and Ord, 1990; Stuart and Ord, 1991)
3. Exponential ordered scores test (Cox and Lewis, 1966).

Luttinen (1996) used directly 3 trend tests for his sampling process, with the exponential ordered scores test as the prime method. Ha et al. (2012) used two sampling methods for headway data sampling. He classified them as short-time and long-time sampling methods. In the short time method, data samples are headway data aggregated with a fixed time interval, 6 minutes in his case. He also used long-time sampling which is the combination of area test validated by three trend tests.

This study uses the combination of area tests and trend tests similar to Ha et al. (2011), provided both velocity and flows are tested. The following steps show the procedure of sampling test:

1. Using area test then trend tests on traffic counts, in the interval of 30 seconds;
2. Using area test then trend tests on velocities, in the interval of 50 counts;
3. Grouping data which have same level of traffic count and velocity;
4. Storing samples that have data points over 500 headway counts.

3.2.2 The Results of Data Sampling

Using the above method, 140 groups of headway samples are achieved. In these samples there are 6 samples, in lane 2, collected on day 1-4 around mid-night, where no headway data available in lane 1. With further checking, those 6 samples are found to have limited velocity around 40 mph, which are unusual traffic flows at those time periods. This raises the suspicion that the road had controlled access and lane 1 was closed. As a result, these 6 samples are removed from this study. 2 samples from lane 0 (the hard shoulder lane) are removed too, since the traffic flows only happens at peak times and the data would not present generality of traffic flows. Table 3.2 shows the number of final selected samples in each of the three lanes.

TABLE 3.2: Number of Samples per lane

	Lane 1	Lane 2	Lane 3
Samples	52	47	33

Figure 3.1 shows how the velocities are distributed against traffic flows for the selected headway samples, categorized by lane number 1-3. Data show widely combinations of velocity and traffic flows in the sampled data, where, lane 2 and 3 have most traffic flow above 1000 veh/hr and more sample in lane 1 have relatively low flow levels below 1000 veh/hr. Lane 3 shows more samples with higher speed, while lane 1 has lower average travelling speed.

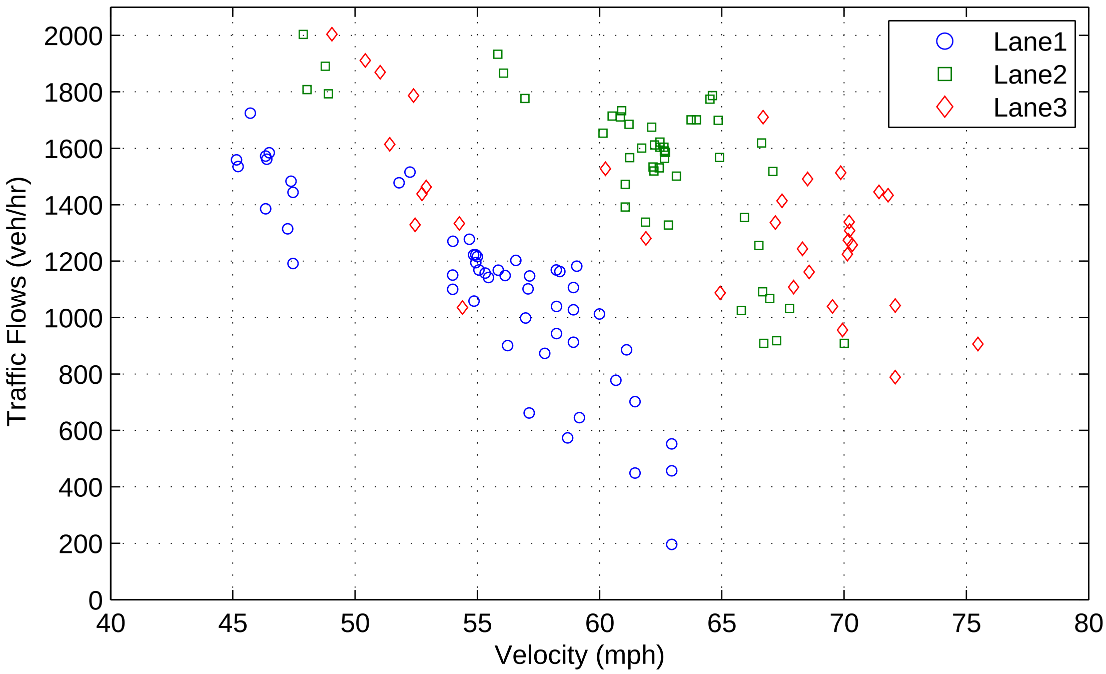

FIGURE 3.2: The Speed and Flow Distributions of Headway Samples

3.2.3 Some Statistics of Data Samples

For the convenience of analysis, the data samples are stored together with the following information and statistics:

1. LaneID, SampleID, Sample Size, Day & Time
2. Mean of flow rates (meanFlow), mean speed(meanSpeed), mean time headway (meanTH), standard deviation of headway (stdTH);
3. Coefficients of variation (CV), skewness, kurtosis, proportion of large good vehicles (pLGV).

Detailed table of data samples are provided in Appendix A, with the above information. Figure 3.3 provides charts that give some general idea about the sampled

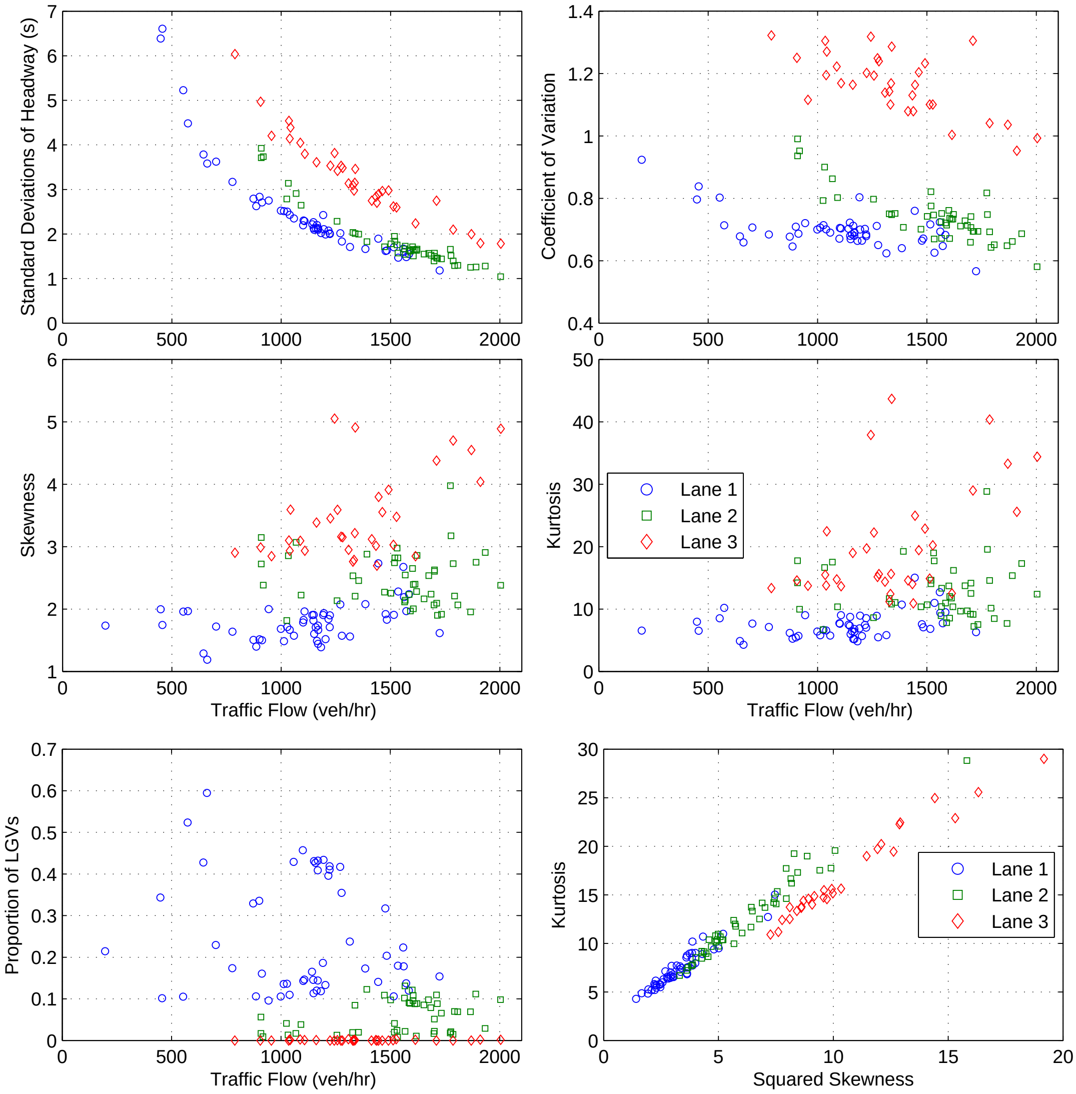

FIGURE 3.3: The Speed and Flow Distributions of Headway Samples

data. The standard deviation of headway data have very close relationships with variation of traffic flows, that is, the values have decreasing trend with increasing flow levels. The coefficient of variation (CV) shows the stability of the headway data, which show more stability of headway in lane 1 and 2 with CV smaller than 1. In lane 3, headway data show more fluctuations with CVs larger than 1, and these fluctuations are more stable while traffic flow gets higher. The lanes 1 and 2 have similar skewness and kurtosis level, with slight growing trend with traffic flows. The data in lane 3 looks more skewed with higher peaks comparing to the rest two lanes. Clear linear relationship are shown in figure (bottom-right) between kurtosis and squared skewness, where the lane 1 and 2 are similar and they are distinguishable from lane 3. Similar data have been reported by Ha et al. (2012) where he used A6a motorway data in France.

The proportion of LGV (large goods vehicles, longer than 6m) data are also presented in Figure 3.3, which shows that most of LGVs are traveling in lane 1. The maximum proportion of LGVs in lane 2 is about 10%, while almost no LGVs travelled in lane 3.

Chapter 4

Mixed Time-Headway Models

Summary

This chapter mainly discusses mixed distribution models used in time-headway analysis. A brief explanation of mixed models and their difference from combined models, is given. Both SPM and GQM mixed models are then discussed.

Second, using *gamma* and *lognormal* distributions as followers headway models, a set of mixed models is investigated with empirical headway samples. In the later sections, methods of simplifying model fittings are attempted by pre-determining one or more parameters. The results are summarized and discussed for each of these attempted methods.

4.1 Mixed Time-Headway Models

As mentioned in chapter 1, many researchers have studied simple distribution models (or single models) to describe headway data, and the main disadvantage is that these simple models are not robust and consistent enough (Ha et al., 2012) in modelling headway data of complex traffic conditions. That is, it is difficult to use just one simple model to fit both the sharp peak and the long tail of the empirical headway distribution (Luttinen, 1996). Most models can only suit one part of the traffic conditions under a limited range of flow rates.

To overcome the limitations of these simple models, many studies have suggested using more than one component in models to describe headway data. In general, these can be termed composite models. Two type of composite models are often discussed, i.e. combined models and mixed models. The majority of these models are composed of two components. Models that contain more than two components, e.g. the study of Ovuworie et al. (1980), have also been suggested. However, due to the complexity of formulation and parameter estimation, these models are not widely used.

4.1.1 Composite Distribution Models

Most studies on composite models suggest separation of vehicles into two groups, i.e. a followers group and a non-followers (also known as leaders) group. In the followers group, all drivers' behaviours are restricted by vehicles in front of them, and they are unable to drive at their desired speed. In contrast, in the non-followers group, vehicles are either leaders of queues or individual vehicles without a follower. The non-followers can drive freely with constraint only related to the overall traffic environment, and no preceding vehicles are immediately ahead of them.

Compared to the non-followers vehicles, the followers group includes some unavoidable approximations that would prevent a composite model from fully corresponding to realistic traffic (Wasielewski, 1979). Luttinen (1996) have summarized 3 or 4 categories of vehicles into the followers group:

1. Followers
2. Vehicles in a transition stage from free-flowing to follower
3. Vehicles starting a passing manoeuvre
4. Vehicles in the stage of passing but still in the original lane would be included as the 4th category.

Vehicles in categories 2, 3 and 4 are assumed as followers as they have relatively small headways which are closer to the headways of true followers.

The common way of studying these two vehicle groups is to composite two probability distributions, which describe the headways of these two groups. Considering headway variable $\mathbf{X}$ as either the followers headway $\mathbf{U}$ or the non-followers headway $\mathbf{V}$, then the distribution of $\mathbf{X}$ can be described as a linear combination of $\mathbf{U}$ and $\mathbf{V}$, as:

$$F(t) = \phi G(t) + (1 - \phi)H(t) \tag{4.1}$$

$$f(t) = \phi g(t) + (1 - \phi)h(t) \tag{4.2}$$

where $F(t), G(t), H(t)$ and $f(t), g(t), h(t)$ are the CDF and PDF of $\mathbf{X}$, $\mathbf{U}$ and $\mathbf{V}$ respectively; the parameter ϕ is the probability of one vehicle being a follower, and it represents the fraction of vehicles in the followers group (Ha et al., 2012).

The above can be used as a generalized form of composite headway models. Based on this, there are mainly two types of composite models that are frequently studied, i.e. combined models and mixed models.

The main difference between combined models and mixed models in the generalized form, is how the non-followers headway variable $\mathbf{V}$ is treated. That is, the methods of deriving $h(t)$ of each type of models are different.

In the combined models, workers use the exponential distribution for the function $h(t)$, as it is natural to consider that the non-followers headway variable $\mathbf{V}$ is a Poisson process. The combined models use a fixed threshold of headway $\mathbf{T}$ to distinguish headways of followers and non-followers groups. That is, for headways that are larger than $\mathbf{T}$, all vehicles will be in the non-followers group, otherwise, they will be followers represented by the distribution function $\mathbf{G}$. In this case, the followers group and non-followers group are completely independent. There are many studies for combined models, and researchers, e.g. (Ha et al., 2012; Luttinen, 1996; Zhang et al., 2007), consider that this type of model is still inadequate in describing headway data. Note that the combined models are NOT studied in this thesis.

4.1.2 Mixed Distribution Models

Mixed models also classify vehicles with two components of followers and non-followers groups, and they have the same generalized form of functions as combined models. However, in contrast to combined models, the two components of mixed models are considered to have mathematical connections between them. Different assumptions of such connections will further distinguish mixed models into different types.

If considering the two components to have a probabilistic relationship between them, then the headway threshold **T**, is not defined anymore as an arbitrarily fixed value. Instead, **T** has been considered as a random variable which has its own distribution. This could lead to a semi-Poisson Model (SPM for short). In contrast, if using a queuing model with Poisson arrivals, and using a general distribution as service time, then the mixed model becomes a generalized queueing model (GQM for short). At present, the SPM and GQM mixed models are probably the most popular models in headway analysis.

4.1.2.1 Buckley's Semi-Poisson Model (SPM)

Buckley (1968) and Wasielewski (1979) considered non-followers as a group of vehicles whose headways come from a Poisson process under the condition that their headways are larger than the random headway threshold **T**. In other words, Buckley suggested that the non-followers headway variable **V**, which is greater than the followers headway **U**, follows an exponential distribution. This random threshold

T in Buckley's model has been considered as a 'zone of emptiness' behind each vehicle which is non-accessible by its following vehicle, in time zone. Within the followers group, this 'zone of emptiness' is exactly the headway of each vehicle behind, which can be described by the distribution function $G(t)$ or density function $g(t)$. Then function $h(t)$ is an exponential density function with parameter λ, modified to include only headways greater than random variable **U**, whose density function is $g(t)$. That is (Buckley, 1968),

$$\begin{aligned} h(t) &= (1/B)\lambda e^{-\lambda t}\int_0^t g(u)du \\ &= (1/B)\lambda e^{-\lambda t}G(t), \qquad\qquad t \geq 0, \lambda \geq 0 \end{aligned} \tag{4.3}$$

where B is the normalization constant:

$$B = \int_0^\infty \lambda e^{-\lambda t}\int_0^t g(u)du\,dt = g^*(\lambda) \tag{4.4}$$

Luttinen (1996) also derived this equation using method of conditional probability, using the relationship between **U** and **V**:

$$\Pr\{V \geq t|V > U\} \sim e^{-\lambda v} \tag{4.5}$$

Buckley's model is named as a *semi-Poisson* model (SPM), with the complete form described as follows:

PDF:

$$f(t) = \phi g(t) + (1-\phi)\frac{G(t)}{g^*(\lambda)}\lambda e^{-\lambda t}, \qquad t \geq 0; \lambda > 0; 0 \leq \phi \leq 1 \tag{4.6}$$

CDF:

$$F(t) = \phi G(t) + \frac{1-\phi}{g^*(\lambda)}\int_0^t G(u)\lambda e^{-\lambda u}\mathrm{d}u, \qquad t \geq 0; \lambda > 0; 0 \leq \phi \leq 1 \tag{4.7}$$

Laplace transform:

$$f^*(s) = \phi g^*(s) + (1-\phi)\frac{\lambda g^*(s+\lambda)}{(s+\lambda)g^*(\lambda)} \tag{4.8}$$

4.1.2.2 The Generalized Queuing Model (GQM)

Considering stability and safety as the modification of Poisson vehicle arrivals is similar to adding on to a Poisson process with a classical queuing system with a single server (Branston, 1976). The followers distribution can be viewed as a service time distribution in the queuing system. Branston suggested using the queuing theory and derived the mixed models into a generalized queuing model (GQM). Various GQMs have been proposed by Cowan (1975) and Branston (1976). Cowan (1975) carried out a similar study of it and named it model M4. In the GQM, non-followers headway variable **V** is proposed as the sum of followers variable **U** and Poisson arrivals **Y**, so that **V**= **U**+**Y**. Then function $h(t)$ becomes the convolution of $g(t)$ and an exponential distribution with parameter λ, it follows that

$$\begin{aligned} h(t) &= \int_0^t \lambda e^{-\lambda(t-u)} g(u) du \\ &= \lambda e^{-\lambda t} \int_0^t g(u) e^{\lambda u} \mathrm{d}u \end{aligned} \tag{4.9}$$

Then the full form becomes:

PDF:

$$f(t) = \phi g(t) + (1-\phi)\lambda e^{-\lambda t} \int_0^t g(u) e^{\lambda u} \mathrm{d}u \tag{4.10}$$

Two ways of deriving CDF:

(i) by Gross and Harris (1985):

$$F(t) = \phi G(t) + (1-\phi) \int_0^t G(t-u) \lambda e^{-\lambda u} \mathrm{d}u \tag{4.11}$$

(ii) equivalently by Luttinen (1996):

$$\begin{aligned} F(t) &= \phi G(t) + (1-\phi) \int_0^t g(u) \big(1 - e^{-\lambda(t-u)}\big) \mathrm{d}u \\ &= G(t) - (1-\phi) e^{-\lambda t} \int_0^t g(u) e^{\lambda u} \mathrm{d}u \end{aligned} \tag{4.12}$$

The Laplace transform of GQM is obtained using the convolution property:

$$\begin{aligned} f^*(s) &= \phi g^*(s) + (1-\phi) g^*(s) h^*(s) \\ &= \phi g^*(s) + (1-\phi) g^*(s) \frac{\lambda}{s+\lambda} \end{aligned} \tag{4.13}$$

From the above, it is clear that the combined models use fixed threshold **T** as the distinction between two headway groups, whereas the mixed models do not. Instead, the mixed models use probabilistic relationships between $g(t)$ and $h(t)$ to describe the non-follower groups. In SPM, the non-followers headways are exponentially distributed under the condition that these headways are greater than followers headways. For GQM, a service time is considered upon Poisson arrivals, hence the non-followers headway is the sum of service time and arrival time. Another difference between combined and mixed models is pointed out by Ha et al. (2012), using the case where parameter $\phi = 0$. That is, when $\phi = 0$, combined models become exponential distributions but mixed models do not.

4.1.3 The Followers Headway Distribution in Mixed Models

In mixed models, the GQM or SPM using the followers headway distribution function, i.e. g function, are called g-GQM or g-SPM respectively (Ha et al., 2012). For example, if a GQM or SPM is using *gamma* distribution as g function, it is called gamma-GQM or gamma-SPM respectively. Some followers headway distributions have been studied in the past with mixed models.

Buckley (1962, 1968) suggested *normal*, *truncated normal* and *gamma* distributions as models of followers headways. At low and low-medium traffic flows, Buckley did not find a solution for parameter estimation of gamma-SPM. However, with parameters that were successfully obtained, he found gamma-SPM had similar or better GoF test result than the rest of the distributions he investigated.

Ashton (1971) attempted a mixed model with the *gamma* distribution because of its connection with the *exponential* distribution. He suggested such a model should be theoretically more tractable.

Cowan (1975) used the *gamma* distribution as service time headway in GQM (M4). He concluded that the M4 model is more realistic for many users and does not necessarily lead to extreme mathematical difficulties.

Branston (1976) applied and compared both SPM and GQM models with *normal*, *gamma* and *lognormal* as followers headway distributions with the GQM model. The *Lognormal* distribution was not applied with SPM as there is no explicit Laplace tranform. He found the results of GoF test were not acceptable when using *normal* distribution for any of SPM or GQM model. He found acceptable and almost identical overall GoF results when applying the *gamma* distribution in both SPM and GQM. Finally, he recommended the lognormal-GQM as it had the best overall fitting result.

Luttinen (1996) tested the gamma-mixed models and found that the gamma-SPM had the best GoF result. He recommended the gamma-SPM for demanding applications with adequate computational facilities.

Zhang et al. (2007) tested the normal-GQM and the gamma-GQM, and found poor GoF results with the headway data he investigated. However, he did suggest that GQM is generally considered more realistic and flexible, and recommended to try more followers headway distribution with GQM model, which quoted as M4 in his paper.

Ha et al. (2012) compared *beta, gamma, normal, log-normal* and *Rayleigh* distributions on both SPM and GQM models. He recommended that the gamma-GQM had the most outstanding result compared with other models. He noted that the gamma-GQM and the gamma-SPM both had similar acceptable GoF results, and he finally suggested that the gamma-GQM was more stable and low-cost (in GoF test time) in overall comparisons.

4.1.4 Using *Gamma* and *Lognormal* Distributions in Mixed Models

In the following sections, the method of using *gamma* and *lognormal* models as followers headway models is selected for further study and test. These two distributions are applied to both GQM and SPM models, that is, gamma-SPM, gamma-GQM, lognormal-SPM and lognormal-GQM.

This section introduces, for each of above mixed models, some mathematical properties, which mainly include PDF, CDF functions and expression of the Laplace transform. For both gamma-SPM and gamma-GQM, more detailed discussions on their mathematical properties can be found in thesis of Luttinen (1996).

The PDF and CDF functions are expressed using simplified forms that are suitable and easier to implement in the proposed framework. As shown in Eq. 4.6 and 4.7, i.e. the general forms of mixed models, if the g functions have been developed as single models, these equations can then directly use them in programming stage, without requiring any further derivations. This feature makes development of mixed models simpler. Of course, some of these will be computationally expensive, as there are complexity of integrations of g functions involved, especially for GQM models, which involve integrations in both PDF and CDF functions. Despite the slow performance, such methods are valid and require minimum effort in the development stage. This will be very useful in testing the fitting quality of a new model in the preliminary stage, without paying too much attention to detailed mathematics.

However, some levels of simplification are always preferred, as long as they do not add great complexities in development (e.g. re-writing a likelihood function), to make implementations run faster. The PDF of SPM is very simple as it only requires the Laplace transform of the g function, and these functions are not very costly in computation. They do not need any further expansion or derivations. Hence, the expressions of PDF in this study, remains in the same form as Eq. 4.6, without further expansion of $g(t)$ or $G(t)$ components.

The CDF of SPM involves the integration of $G(t)$ for non-followers headway, and this will normally cause slow performance on GoF test stage, or on random number generation. The cost of the GoF test does not sound too problematic, as the demands (number of iterations involved) of computation are trivial compared to parameter estimation. However, in some applications such as random number generation using CDF function, the computational cost would be unbearable, if there is not an alternative method available. With the $G(t)$ involved in integration, it is more difficult to simplify a CDF of SPM to obtain more explicit forms. In this study, if there is no simplification of CDFs, they will remain the same form as Eq. 4.7, otherwise, the simplified version will be given for improvement.

For GQM mixed models, both PDF and CDF (Eq. 4.10 and 4.12) involve integrations of $g(t)$ times an exponential component $e^{\lambda t}$. These integrations again will affect the computational performance, in the stage of parameter estimation, GoF test etc. If no explicit form of these functions is available, e.g. lognormal-GQM (Ha et al., 2012), implementation of these models will be unbearably slow. Some of these functions can be derived with explicit forms, in which case their expressions will be given.

Note that lognormal-SPM is not studied in any other papers as no closed form of Laplace transform is available for *lognormal* distribution. However, since Asmussen et al. (2014) has proposed an approximation method for Laplace transform of *lognormal* distribution, the implementation of lognormal-SPM becomes possible. Hence, the performance of this model is also investigated in this study.

4.1.4.1 Gamma-SPM

The PDF and CDF of *gamma* distribution are given in Eq. 2.15 and 2.17, which is denoted here as $g_g(t)$ and $G_g(t)$ respectively. Its Laplace transform Eq. 2.21 is denoted as $g_g^*(s)$. Substituting $g_g(t)$, $G_g(t)$ and $g_g^*(\lambda)$ in Eq. 4.6, will gain PDF of gamma-SPM model, which is:

$$f(t;\phi,\alpha,\beta,\lambda) = \phi\, g_g(t;\alpha,\beta) + (1-\phi)\frac{G_g(t;\alpha,\beta)}{g_g^*(\lambda)}\lambda e^{-\lambda t} \tag{4.14}$$

Accordingly, substituting $G_g(t)$ and $g_g^*(\lambda)$ in Eq. 4.7, will give the CDF of gamma-SPM model, which is:

$$F(t;\phi,\alpha,\beta,\lambda) = \phi\, G_g(t;\alpha,\beta) + \frac{1-\phi}{g_g^*(\lambda)}\int_0^t G_g(u;\alpha,\beta)\lambda e^{-\lambda u}\mathrm{d}u \tag{4.15}$$

As mentioned above, the CDF function involves integration of $G_g(t)$, this slows down the performance when GoF test is needed. This CDF can be further expanded as (Luttinen, 1996):

$$F(t;\phi,\alpha,\beta,\lambda) = \phi\frac{\gamma(\alpha,\beta t)}{\Gamma(\alpha)} + (1-\phi)\frac{\lambda}{\Gamma(\alpha)}(1+\frac{\lambda}{\beta})^{\alpha}\int_0^t \gamma(\alpha,\beta u)e^{-\lambda u}\mathrm{d}u \tag{4.16}$$

where $\gamma(.)$ is the incomplete gamma function, cf. Eq. 2.18.

In Eq. 4.16, its integration part can be further derived as:

$$\int_0^t \gamma(\alpha,\beta u)e^{-\lambda u}\mathrm{d}u = \frac{1}{\lambda}\left[\left(\frac{\beta}{\lambda+\beta}\right)^{\alpha}\gamma\left(\alpha,(\lambda+\beta)t\right) - \gamma(\alpha,\beta t)\, e^{-\lambda t}\right] \tag{4.17}$$

after substitution and simplification, Eq: 4.16 becomes

$$F(t;\phi,\alpha,\beta,\lambda) = \phi\, G_g(t;\alpha,\beta) + (1-\phi)\left[G_g(t;\alpha,\lambda+\beta) - \frac{G_g(t;\alpha,\beta)}{g_g^*(\lambda)}e^{-\lambda t}\right] \tag{4.18}$$

The Laplace transform of gamma-SPM model (Luttinen, 1996):

$$f^*(s) = \phi\left(\frac{\beta}{\beta+s}\right)^{\alpha} + (1-\phi)\frac{\lambda}{\lambda+s}\left(\frac{\beta+s}{\beta+\lambda+s}\right)^{\alpha} \tag{4.19}$$

4.1.4.2 Gamma-GQM

Similar to the substitutions in the derivation of the gamma-SPM, substituting $g_g(t)$ and $G_g(t)$ into Eq. 4.10, will give the the PDF and CDF of the gamma-GQM model respectively, which are:

PDF of gamma-GQM:

$$f(t;\phi,\alpha,\beta,\lambda) = \phi\, g_g(t;\alpha,\beta) + (1-\phi)\lambda e^{-\lambda t}\int_0^t g_g(u;\alpha,\beta)e^{\lambda u}\mathrm{d}u \tag{4.20}$$

which is equivalent to (Luttinen, 1996):

$$f(t;\phi,\alpha,\beta,\lambda) = \phi g_g(t;\alpha,\beta) + (1-\phi)\left(\frac{\beta}{\beta-\lambda}\right)^{\alpha} G_g(t;\alpha,\beta-\lambda)\,\lambda e^{-\lambda t} \tag{4.21}$$

referring to Eq. 2.21, since

$$\left(\frac{\beta}{\beta-\lambda}\right)^{\alpha} = g_g^*(-\lambda)\ .$$

The PDF can be also expressed as

$$f(t;\phi,\alpha,\beta,\lambda) = \phi g_g(t;\alpha,\beta) + (1-\phi)\,\lambda g_g^*(-\lambda) G_g(t;\alpha,\beta-\lambda)\,e^{-\lambda t} \qquad t\geq 0; \lambda>0; 0\leq\phi\leq 1 \tag{4.22}$$

Accordingly, substituting $g_g(t)$ and $G_g(t)$ into Eq. 4.12, will give the CDF of the gamma-GQM model, which is:

$$F(t;\phi,\alpha,\beta,\lambda) = G_g(t;\alpha,\beta) - (1-\phi)e^{-\lambda t}\int_0^t g_g(u;\alpha,\beta)e^{\lambda u}\mathrm{d}u \tag{4.23}$$

which is equivalent to (Luttinen, 1996):

$$F(t;\phi,\alpha,\beta,\lambda) = G_g(t;\alpha,\beta) - (1-\phi)\left(\frac{\beta}{\beta-\lambda}\right)^{\alpha} G_g(t;\alpha,\beta-\lambda)\, e^{-\lambda t} \tag{4.24}$$

Again, it can be rewritten as

$$F(t;\phi,\alpha,\beta,\lambda) = G_g(t;\alpha,\beta) - (1-\phi)\, g_g^*(-\lambda)G_g(t;\alpha,\beta-\lambda)\, e^{-\lambda t} \tag{4.25}$$

The Laplace transform of gamma-GQM (Luttinen, 1996):

$$f^*(s) = \left(\frac{\beta}{\beta+s}\right)^{\alpha}\left(\phi + (1-\phi)\frac{\lambda}{\lambda+s}\right) \tag{4.26}$$

4.1.4.3 Laplace Transform of Lognormal Distribution

The PDF and CDF of the *lognormal* distributions are introduced in section 2.4.2. Its Laplace transform is introduced here. The Laplace transform is important as it forms a part of the SPM mixed models.

The Laplace transform:

$$\begin{aligned} g^*(\lambda) &= \int_0^\infty e^{-\lambda x} \mathrm{d}F_{\mu,\sigma^2}(x) \\ &= \int_0^\infty \frac{1}{x\sqrt{2\pi}\sigma} \exp\left\{-\lambda x - \frac{(\log x - \mu)^2}{2\sigma^2}\right\} \mathrm{d}x \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\lambda e^t - \frac{(t-\mu)^2}{2\sigma^2}\right\} \mathrm{d}t \end{aligned} \tag{4.27}$$

This function cannot be further derived, and the the explicit form of Laplace transform does not exist. Recently, a method of approximation was proposed by Asmussen et al. (2014).

$$\tilde{g}^*(\lambda) = \frac{\exp\left\{-\frac{W^2(\lambda e^\mu \sigma^2) + 2W(\lambda e^\mu \sigma^2)}{2\sigma^2}\right\}}{\sqrt{1 + W(\lambda e^\mu \sigma^2)}}, \lambda \in \mathbb{R}^+ \tag{4.28}$$

In the expression above, $W(\cdot)$ is the Lambert W function which is defined as the solution of the equation $W(x)e^{W(x)} = x$; this function has been widely studied in the last 20 years mainly due to the advent of fast computational methods (Corless et al., 1996).

4.1.4.4 Lognormal-SPM

With the approximation of Laplace transform available, the *lognormal* distribution can be now considered in SPM model. Similar to procedures for gamma-SPM, by substituting Eq. 2.24, 2.25 and 4.28 into Eq. 4.6 and 4.7, approximations of the PDF and CDF can be achieved for the lognormal-SPM model, which are:

PDF:

$$\tilde{f}(t;\phi,\mu,\sigma,\lambda) = \phi\, G_l(t;\mu,\sigma) + \frac{1-\phi}{\tilde{g_l}^*(\lambda)} G_l(t;\mu,\sigma)\lambda e^{-\lambda t} \tag{4.29}$$

CDF:

$$\tilde{F}(t;\phi,\mu,\sigma,\lambda) = \phi G_l(t;\mu,\sigma) + \frac{1-\phi}{\tilde{g_l}^*(\lambda)} \int_0^t G_l(u;\mu,\sigma)\lambda e^{-\lambda u}\mathrm{d}u \tag{4.30}$$

4.1.4.5 Lognormal-GQM

Both PDF and CDF of lognormal-GQM can be achieved, by substituting Eq. 2.24 and 2.25 into Eq. 4.10 and 4.12, as addressed below:

PDF:

$$f(t;\phi,\mu,\sigma,\lambda) = \phi g_l(t;\mu,\sigma) + (1-\phi)\lambda e^{-\lambda t} \int_0^t g_l(u;\mu,\sigma)e^{\lambda u}\mathrm{d}u \tag{4.31}$$

CDF:

$$F(t;\phi,\mu,\sigma,\lambda) = G_l(t;\mu,\sigma) - (1-\phi)e^{-\lambda t} \int_0^t g_l(u;\mu,\sigma)e^{\lambda u}\mathrm{d}u \tag{4.32}$$

Unfortunately, it is more difficult to simplify the above functions for both lognormal-SPM and lognormal-GQM models, and they have to be left with involvement of integrations. These will greatly reduce speed of implementation for both models. Also, no explicit form of Laplace transform exists, for obvious reasons.

4.2 Implementing Mixed Models in MATLAB

The methods of implementing single models using MATLAB for headway analysis are introduced in Chapter 2. In this section, a further extension for adding mixed models into the proposed framework is explained. Using the advantages of the MATLAB Statistical Toolbox and object oriented programming structure, a simple method is introduced that can provide the ability to quickly implement new followers headway models into the GQM or SPM mixed model in MATLAB. This gives the researcher more advantages of testing new models in a quick and reliable way, and saves substantial development time and effort in the preliminary research stage.

4.2.1 Method to add mixed models in MATLAB

The data structure and class structure of adding new models in the Statistics Toolbox have been explained in chapter 2. Here, attention is more focused on a new class structure that can treat the mixed models in a similar way as other simple models. Since the mixed model mainly has two components as described in Eq. 4.2, it would be an advantage to use the similarity of this common form of models to simplify the development work. This proposed attempt is to minimize the work by avoiding to write similar codes for every mixed model.

The SPM and GQM are treated separately in this method, as there are fundamental differences between methods of dealing with non-followers distribution, e.g.

the $h(t)$ and $H(t)$. This would not cause many complications as there are only two parent classes resulting in the end, i.e. one for each type of the two models.

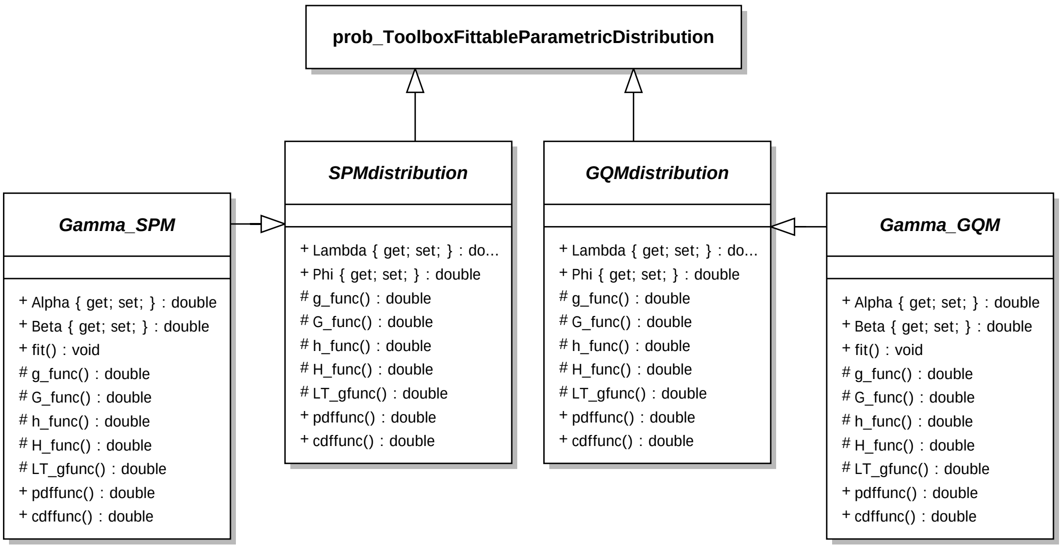

FIGURE 4.1: Class diagram of mixed models, implemented in proposed framework

The class structure is shown in Figure 4.1 using gamma-Mixed models as an example. The two parent classes, i.e. the *SPMdistribution* and *GQMdistribution*, are also subclass of `prob.ToolboxFitableParametricDistribution` just like any of the single models in chapter 2. Only that these two are abstract classes that cannot be instantiated directly.

The details of properties and methods are not explained here as it would take large amount of space, and similar content can be found from any of the Statistics Toolbox documentation. The added properties are `Lambda` and `Phi`, which are representing λ and ϕ in Eq. 4.6 and 4.10. Five new methods are added in these two classes, which are `g_func, G_func, h_func, H_func and LT_g_func`, and these methods correspond to $g(t), G(t), h(t)$ and $H(t)$ functions in Eq. 4.2. The `LT_g_func` is Laplace transform of $g(t)$ for followers headway. These two classes also override methods of `pdffunc` and `cdffunc`, to represent the PDF and CDF of mixed models, by implementing Eq. 4.2. Methods `g_func`, `G_func` and `LT_g_func` are abstract methods, which have to be overridden in subclasses with the actual PDF and CDF code of followers headway models.

Both classes `SPMdistribution` and `GQMdistribution` are very similar in structure, in fact they are almost identical, except methods `h_func` and `H_func`, as functions $h(t)$ and $H(t)$ are treated differently for SPM and GQM models. These two methods are implemented corresponding to the common form of functions of SPM (cf. Eq. 4.6 and 4.7) and GQM (cf. Eq. 4.6 and 4.12). These two classes are designed in such a way so `h_func` and `H_func` methods don't have to be overridden, and their inherited subclass can still be functional, as long as the follower headway distributions have been developed. Such design will help greatly if a researcher only needs to have a quick check on how a single model will perform on headway data, when it is used as part of a mixed model, without going through detailed mathematics for simplification and optimization. Of course, methods of such will be computationally slow, for obvious reasons. An example in this chapter, the lognormal-mixed models do not have explicit functions, and this method is used for their implementations.

Using the *gamma* distribution as an example, as shown in the class diagram, `Gamma_SPM` and `Gamma_GQM` are subclass of `SPMdistribution` and `GQMdistribution` respectively. To let these two subclasses functional, firstly their methods of `g_func`, ... `G_func` and `LT_g_func` have to be overridden. Since the *gamma* distribution has been provided by Matlab, its methods can be directly called. For instance, the `g_func` can be as following:

```
function y = g_func(x,varargin)      % PDF g(t) of followers ... headway distribuiton
     pars = [varargin{:}];
     a = pars(1);
     b = pars(2);
     y = prob.GammaDistribution.pdffunc(x,a,b); %PDF of gamma ... distribution
end
```

and this will work for both gamma-SPM and gamma-GQM models. Since the simplified forms of functions, e.g, Eq. 4.18 for CDF of gamma-SPM, the `H_func` can be overridden in this case for `Gamma_SPM` subclass, and this will greatly improve its computational performance.

After the subclass of a mixed model is developed, it is ready for fitting headway data, with numerical MLE methods, and this is the way applied for all mixed models throughout this study. If there are further simplified forms of PDF or CDF available, and they are required to be used in the model fitting, then these new expressions can be coded by overriding the methods `pdffunc` and `cdffunc`, which will overwrite the common form of functions with the more specific ones developed. Furthermore, if alternative estimation methods are required, for instance, a good solution for a likelihood function, or a MCS, they can be used by overriding the `fit` method.

Overall, the above section introduced how to add new mixed models to the proposed framework, and this method is used for implementation of all mixed models involved in this study.

4.2.2 Parameter Estimation

For mixed distribution models, parameter estimation becomes more difficult due to the more complex structures of their PDFs.

Buckley (1968) used numerical methods on the moment estimators (ME) to estimate the parameters of SPM models. He estimated normal-SPM and gamma-SPM, and he noticed that some of the samples with low and low-medium flow levels were unable to obtain any estimation results. Ashton (1971) estimated a modified gamma-SPM model by minimizing the sum of the squared errors between analytical and calculated moments, which is actually a modified moment estimator. Wasielewski (1974, 1979) suggested a method to examine the tails of the empirical headway distribution, and hence obtain estimations for λ and ϕ. Hoogendoorn (2005) also used and recommended this method in his study of free speed distribution. Branston (1976) used the combination of ME and MCS (minimum of chi-squared statistic), for both SPM and GQM models. For estimation of λ and ϕ, he used the first two moments to equate the sample moment and solved the parameter values. Then he used MCS to estimate α and β for the gamma-mixed models. Luttinen (1996) suggested the use of the MLE method in estimating the gamma-SPM in his study. Hoogendoorn and Botma (1997) proposed a method to estimate GQM parameters by minimizing the mean integrated squared error (MISE) distance in the frequency domain. His method,in many cases, has been proved powerful and not complicated in calculation. Zhang et al. (2007) suggested a method of minimizing the sum of the squared errors (MSSE) between the estimated and empirical distribution functions for estimation of the mixed models he studied. Ha et al. (2012) reviewed and evaluated some of the above estimation methods (MCS, ME, MLE), and he concluded with the recommendation to use MLE accompanied with his proposed three-step estimation process.

The parameter estimation in this study uses the numerical MLE method built into the proposed framework, as discussed in chapter 2. This method does not give the best performance nor the most reliable way to estimate parameters for mixed distribution models, but the overall performance is acceptable, plus its convenience of implementation in the proposed framework. This estimation method can be considered as an attractive alternative if no better method has been identified.

4.2.3 Fitting Mixed Models Using Empirical Headway Samples

Once the gamma-mixed and the lognormal-mixed models are implemented using the methods explained in previous section, application of these mixed models on headway data are the same as those of using single models discussed in chapter 2. Each model will be used to fit all the headway samples described in chapter 3 and then estimated models are tested for the quality of fitting using both KS and AD GoF tests. During the fitting and the GoF test stages, all the elapsed cpu-time[1] and the GOF results are recorded for later comparison. These two are the major measurement of comparing the performance of mixed models. Results are listed in data tables as well as figures, for discussions of fitting and the GoF test speeds,

[1]Such elapsed cpu-time is also called GoF time in this thesis. It is the time taken to implement a GoF test, and it is measured in the same way as estimation time, in milliseconds.

quality of fitting and patterns and variations of estimated parameters. Also, PDF curves of a few selected samples are presented to compare with the histograms of headway data, to give visual examinations of the fitting results.

In the parameter estimation stage, the numerical MLE requires initial values for iterative maximization algorithm used. These values are crucial, as poor values may lead to bad fitting results. Two considerations are used for set the initial values. First, for the common parameters ϕ and λ, as scopes of both values are effectively between 0 and 1, so a value of 0.5 is chosen as starting point of these two parameters. Second, for parameters that are specific to followers headway models, they are chosen with values near the estimated values when corresponding models were fitted as single models. For example, when the *gamma* distribution was implemented in chapter 2, the mean estimated results of α and β are used as the starting points of fitting gamma-mixed models. After the first implementation, new mean results will be estimated and they are then selected as final choice of initial values. In practice, more combinations of initial values were tested, especially for gamma-mixed models, and resulted parameters have shown good consistency for these models.

4.2.4 Results and Discussion

All chosen samples are tested in this stage, and in general, the test results are well accepted. All headway samples are categorized with road lane 1-3 as discussed in chapter 3. The gamma-mixed models are generally processed with good estimation time, where each estimation took roughly between 0.5 to 3 seconds. However,

Table 4.1: Comparison of estimation time(s) on gamma-SPM, gamma-GQM, lognormal-SPM and lognormal-GQM models

	Fitting Time(s)		KS test time(ms)		AD test time(ms)	
	Mean	Std	Mean	Std	Mean	Std
Gamma-						
SPM	1.52	1.41	7.96	6.80	5.61	2.46
GQM	1.05	0.61	4.43	2.71	3.27	1.11
Lognormal-						
SPM	5.92	1.88	368.52	119.03	385.18	155.40
GQM	96.05	43.08	373.47	137.62	393.54	176.18

the average speeds on fitting the *lognormal* followers headway models are much slower, which are expected and discussed earlier. For the lognormal-SPM, each sample took nearly 10 seconds to implement. This is due to the relative low performance on the numerical process involved in the approximated Laplace transform when computing the PDF. The fittings of the lognormal-GQM to headway samples are much slower, as each sample took about 150-400 seconds. This slow performance is mainly caused by the unavoidable numerical integrations involved in the computation of the PDF. In practical terms, this may be improved by using alternative optimization methods, such as the Trust-Region algorithm, which would potentially reduce the number of iterations involved in the MLE process. Other estimation methods such as the MCS (Branston, 1976) may also be alternatives for improving the estimation speed. Table 4.1 also provides the time spent on the GoF tests. The gamma-SPM model is very fast in the GoF test with time average lower than 10 milliseconds, and the gamma-GQM is slightly faster than the gamma-SPM. The GoF tests with the lognormal-mixed models are much slower, with the average test time between 200-500 milliseconds. Again, such slow performance is caused by the numerical integrations involved in evaluating CDF of lognormal-mixed models. Ha et al. (2012) reported a much slower performance of the GoF test with the gamma-SPM, comparing with the gamma-GQM, which is one of two reasons he rejected the gamma-SPM model. This may be caused by the involvement of numerical integrations in CDF computation, as he may not have used any explicit solution. However, this has been improved greatly here by using Eq. 4.18.

Table 4.2 shows a summary GoF test result categorized by lane numbers 1- 3. More detailed results are shown in Figure 4.2 and 4.3 for gamma-mixed and lognormal-mixed models respectively.

Table 4.2: Summary of goodness-of-fit results on gamma-SPM, gamma-GQM, lognormal-SPM and lognormal-GQM models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
Gamma-						
SPM	50(2)	52(0)	47(0)	47(0)	33(0)	33(0)
GQM	50(2)	52(0)	47(0)	47(0)	33(0)	33(0)
Lognormal-						
SPM	51(1)	52(0)	47(0)	47(0)	31(2)	31(2)
GQM	52(0)	52(0)	47(0)	47(0)	32(1)	32(1)

In general, all the test results seem reasonably similar, with only 1-2 out of 132 samples not passing the tests. These results show the excellent fitting qualities of all mixed models tested.

Referring to the details of the results on each of the samples, the test statistics of KS and AD GoF are very close to the gamma-SPM and gamma-GQM, in most

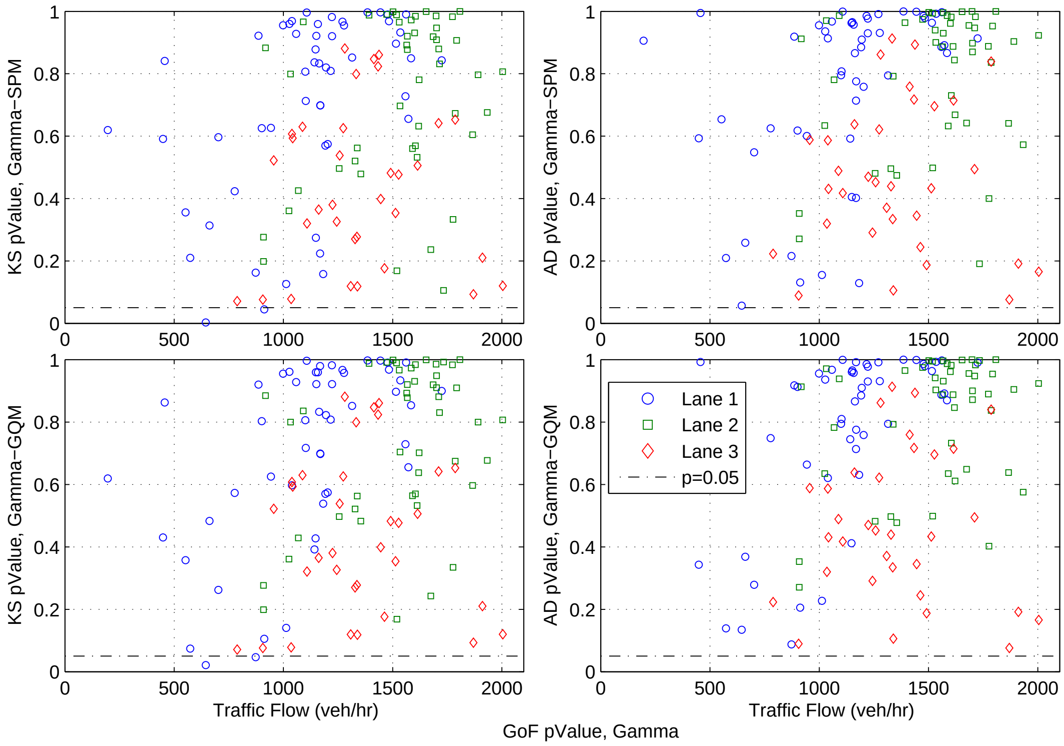

FIGURE 4.2: GoF result (p value) for gamma-SPM and gamma-GQM

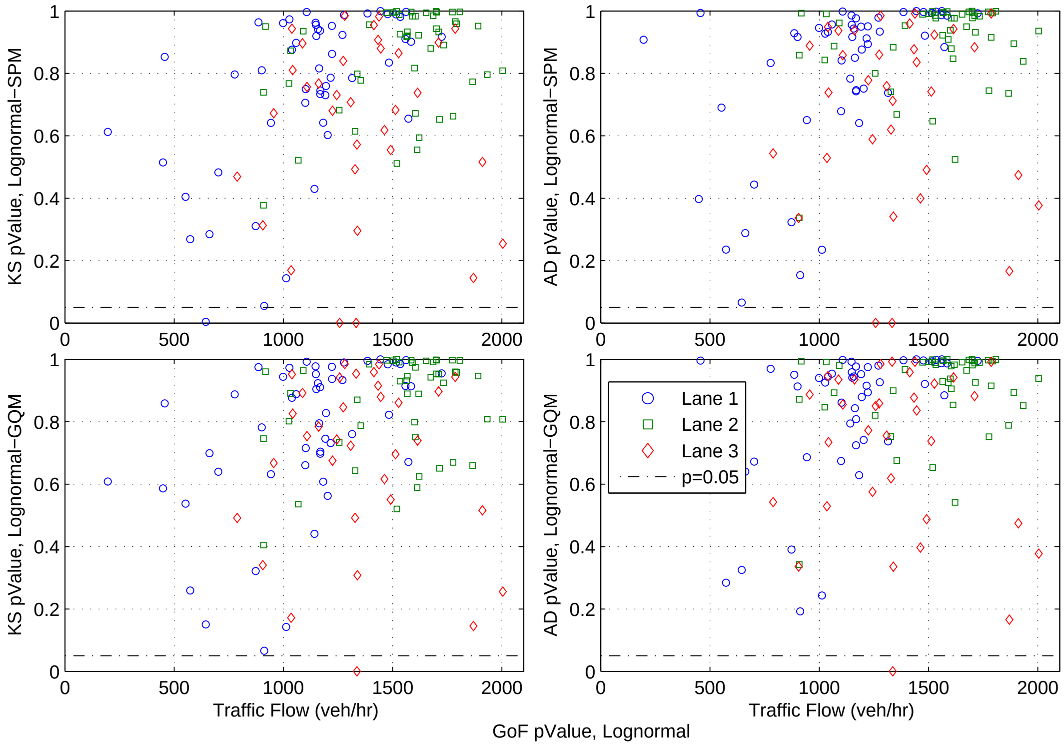

FIGURE 4.3: GoF result (p value) for lognormal-SPM and lognormal-GQM

of the cases. This indicates that the two models of SPM and GQM have similar fitting quality. Similar results were reported by Branston (1976), Luttinen (1996) and Ha et al. (2012). There are slight differences in the findings of Ha et al. (2012), as he found only the KS test results are similar between the two gamma-mixed models. He pointed out that, with some of the samples, the AD statistics of the gamma-SPM are very high comparing to the gamma-GQM model. He used this fact as the second weakness points to reject the gamma-SPM, apart from the concerns of lower performance on the GoF time mentioned earlier. However, with this study, no such significant differences are found that would distinguish between these two models. These two objections either have been improved or have not been observed in this study, and thus the performance of both gamma-mixed models should be considered similar. The lognormal-SPM is tested for the first time, and comparison of KS and AD statistics are also very close to the two lognormal-mixed models. This is further evidence that the similarity of the SPM and GQM does not depend on the type of followers headway models being used.

Figure 4.4 shows visual fitting results for some selected samples, using the PDF of each estimation to compare with the histogram of the sample headway data. The three rows of figures present samples from the road lane 1-3, from top to bottom. The left column shows the samples with relatively low flow rates, while the right column shows the samples with high flow rates. The actually fitting results and GoF test results can be found in tables of Appendix B. Some basic sample information is shown within the figures.

All the figures include the PDF curves of all four mixed models. In these figures,

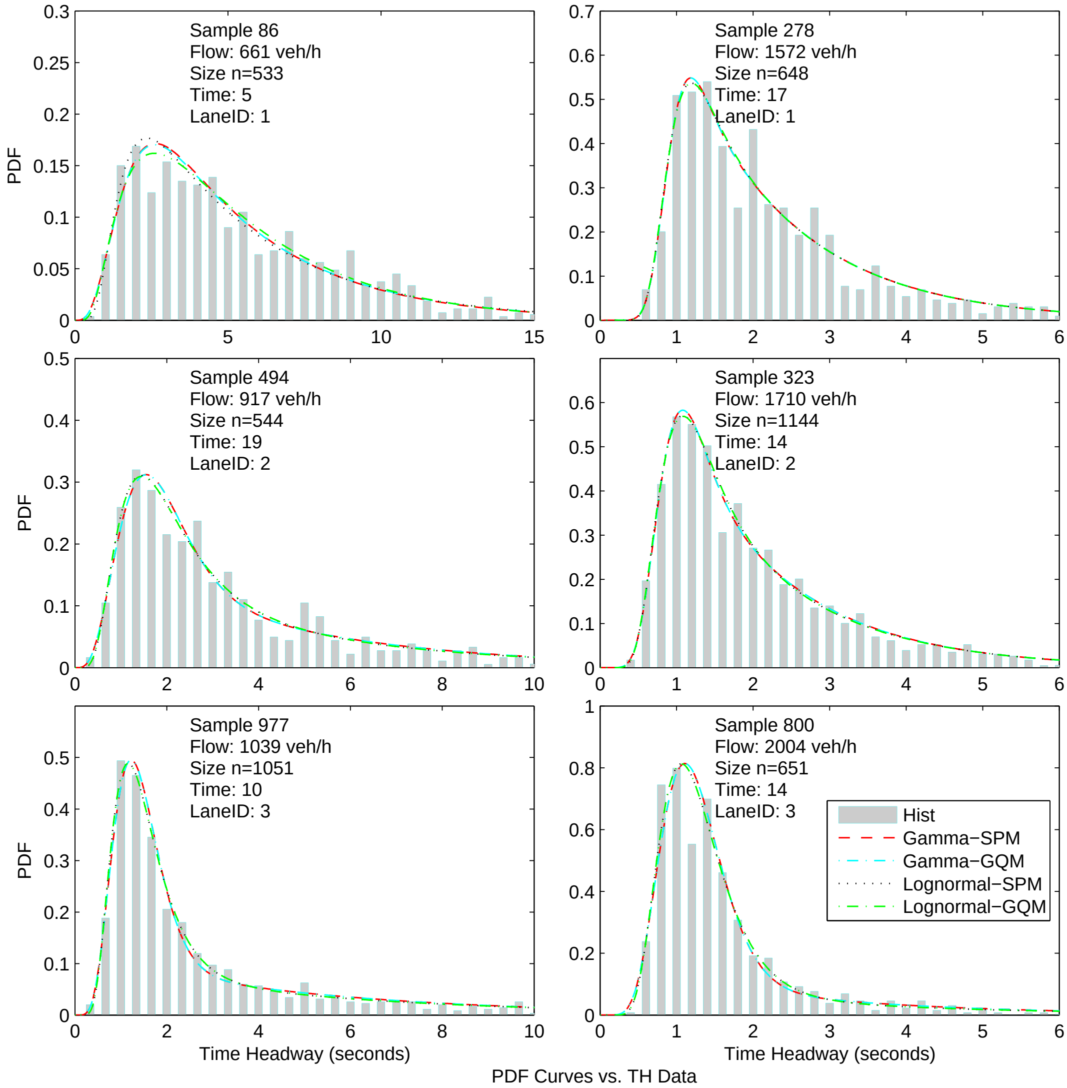

FIGURE 4.4: PDF of selected headway samples

the gamma-mixed models and lognormal-mixed models have only a slight difference between curves. It is hard to see any difference between the SPM and GQM type of models. They are nearly overlapped with each other, and they give good visual fits to the headway data.

Overall from the data and figures, all the mixed models seem to fit the headway samples very well. However, the estimation time would certainly give some distinctions, as the lognormal-mixed models are slower to fit, especially the lognormal-GQM. The lognormal-GQM is well tested by Branston (1976), who reported acceptable results for all the samples he used. As a result of this study, the lognormal-GQM does not seem outstanding compared with the two gamma-mixed models. The lognormal-GQM was also tested by Ha et al. (2012), who suggested that the model did not give good results with the use of the MLE method, and it was exorbitantly slow when using MCS methods to estimate parameters. Luttinen (1996) addressed that the parameter estimation of GQM is problematic.

Considering that Branston (1976) was not able to implement to lognormal-SPM to compare with GQM model, this part of study can be considered as an valuable extension of Branston's work with the *lognormal* as followers headway distribution.

4.2.5 Parameter Variation of the Mixed Models

Table 4.3 and Figures 4.5-4.8 show the estimated parameters across all the four mixed models and for those headway samples that have been accepted by GoF tests. All the figures indicate how the parameters vary with the corresponding traffic flow. In those figures, the samples are categorized into 3 lane groups as mentioned in the previous section.

In all four models, the vehicle arrival rates λ have clear trends as the traffic flow increases. This certainly supports the fundamental concept of constructing the SPM and GQM mixed models. It is worth looking at the difference between the

TABLE 4.3: Parameters variation on gamma- and lognormal-mixed models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Gamma-SPM						
ϕ	0.138	0.152	0.272	0.154	0.567	0.083
λ	0.441	0.145	0.586	0.145	0.287	0.065
α	9.345	4.287	10.228	2.581	9.096	1.948
$1/\beta$	0.186	0.145	0.124	0.058	0.138	0.040
Gamma-GQM						
ϕ	0.178	0.161	0.312	0.155	0.573	0.084
λ	0.438	0.143	0.585	0.144	0.287	0.065
α	10.529	4.655	10.288	2.737	9.098	1.948
$1/\beta$	0.154	0.089	0.126	0.058	0.138	0.040
Lognormal-SPM						
ϕ	0.22	0.20	0.42	0.17	0.62	0.09
λ	0.43	0.14	0.54	0.15	0.27	0.06
μ	0.30	0.29	0.23	0.18	0.15	0.10
σ	0.36	0.10	0.39	0.07	0.38	0.05
Lognormal-GQM						
ϕ	0.28	0.17	0.49	0.16	0.62	0.09
λ	0.43	0.13	0.54	0.14	0.27	0.06
μ	0.37	0.23	0.25	0.18	0.15	0.10
σ	0.38	0.09	0.40	0.07	0.37	0.05

three lane groups where lanes 1 and 2 have very similar trends, with the λ of lane 2 slightly lower than lane 1, while arrival rates of lane 3 seem significantly lower compared with the first two lanes. The mean values of this parameter in Table 4.3 also support this difference. On all the 3 lanes, the λ values seem more scattered in the higher flow of traffic, and it may relate to the increasing fluctuations with heavier traffics.

Between the two gamma-mixed models, the estimated parameters α and β[1] look similar, where α all locate near values between 5-15 and the reciprocals of βlocate near values between 0.05- 0.2. These two parameters do not seem to vary with

[1]Tables, including later ones in chapter 4, use $1/\beta$, as this reciprocal form is used in PDF of *gamma* distribution, provided by Matlab. Note that, β is still used in main text for convenience.

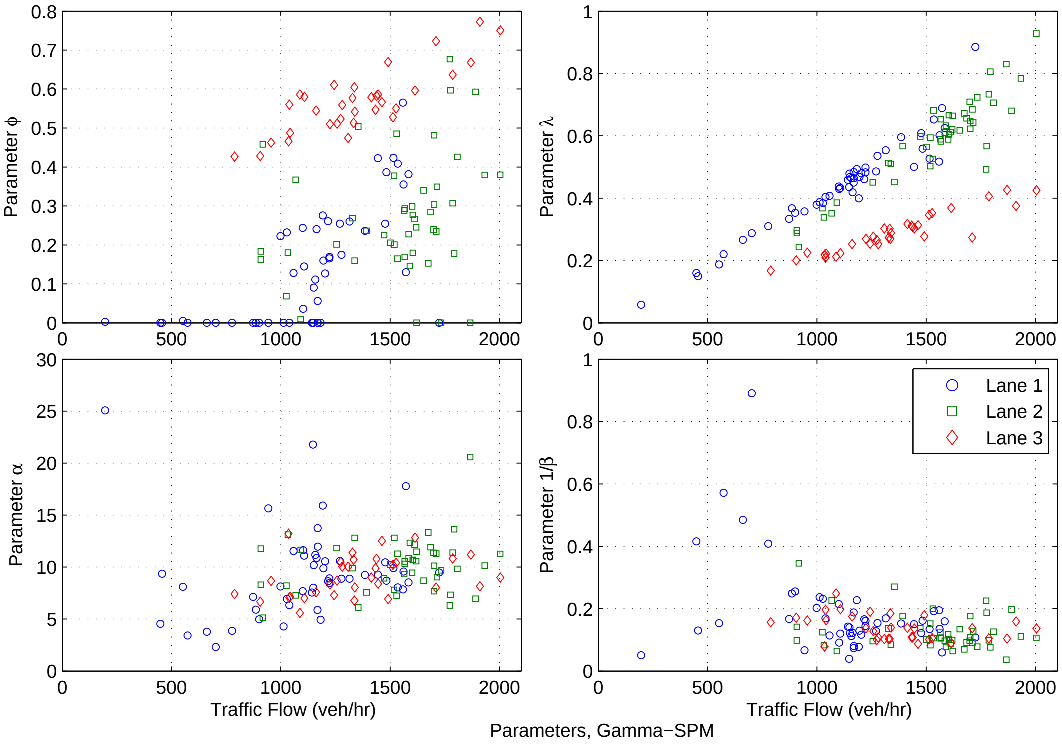

FIGURE 4.5: Estimated parameters for gamma-SPM

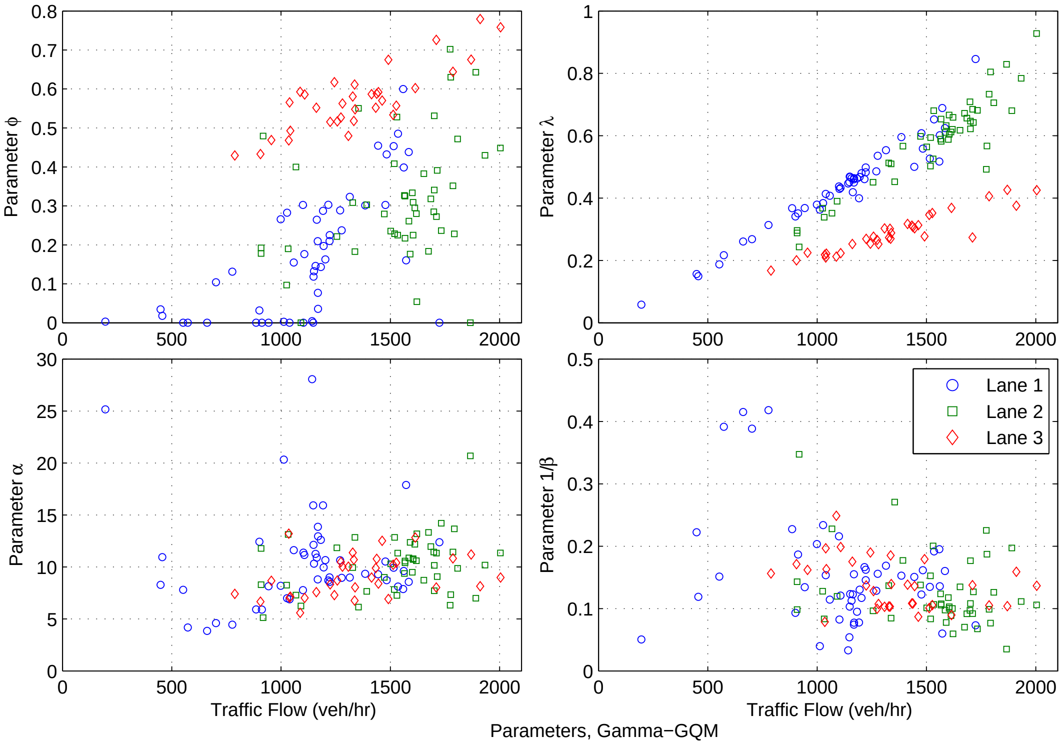

FIGURE 4.6: Estimated parameters for gamma-GQM

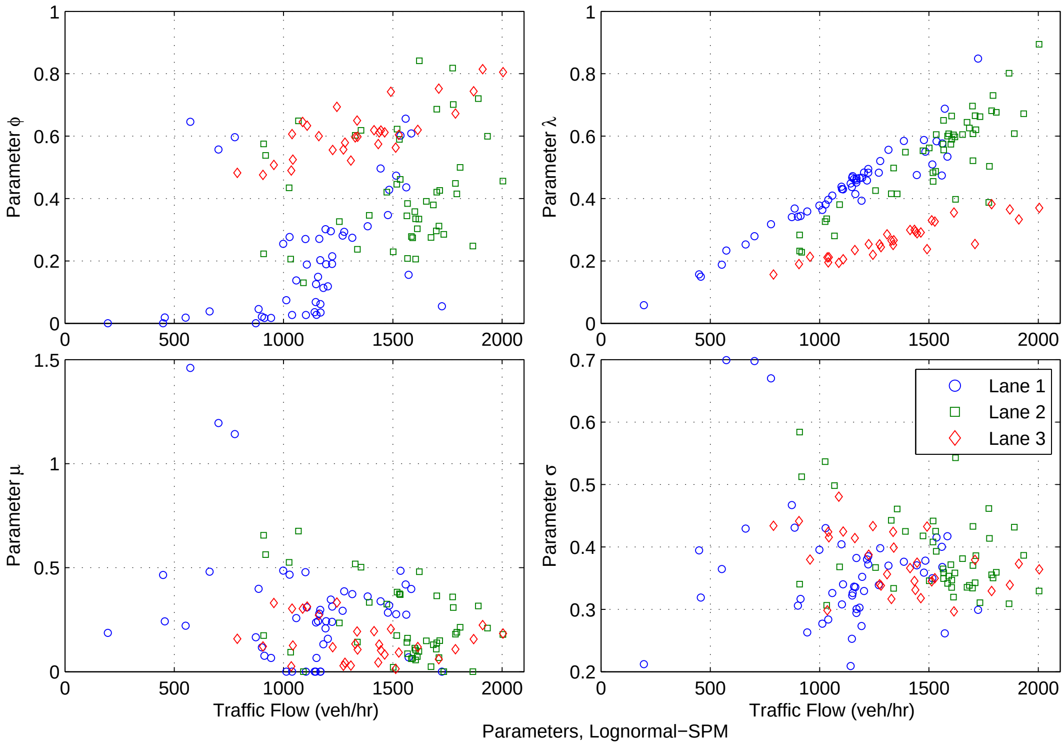

FIGURE 4.7: Estimated parameters for lognormal-SPM

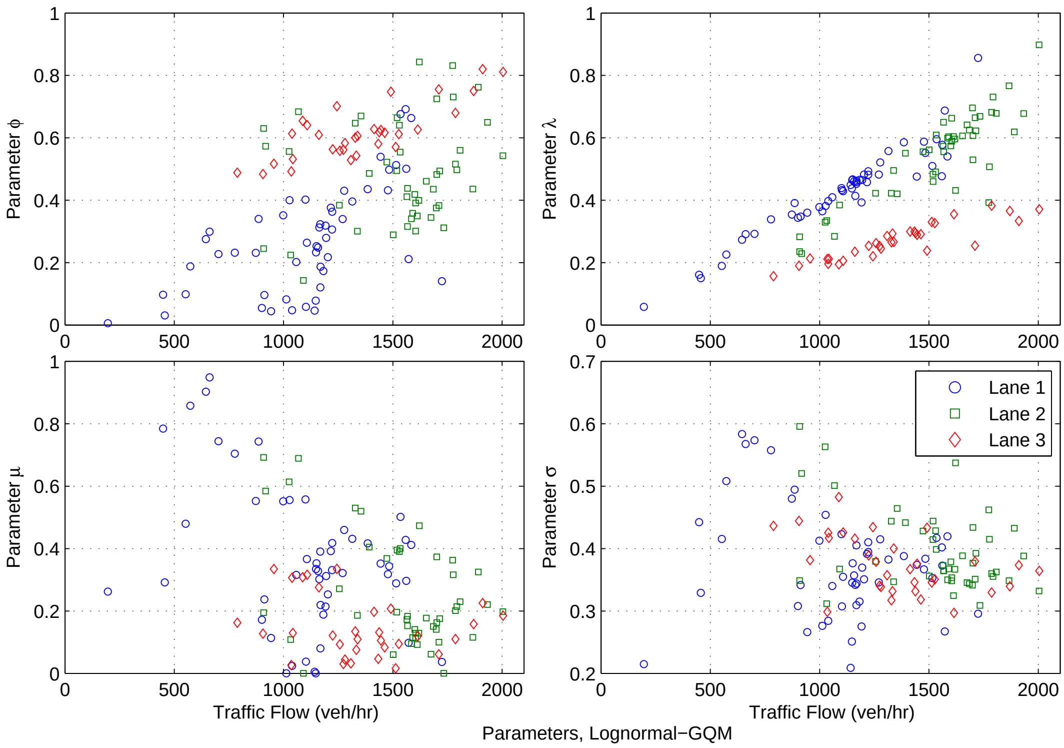

FIGURE 4.8: Estimated parameters for lognormal-GQM

the increased traffic flows, and they do not seem to deviate across the three lanes. At low traffic flow, both α and β seem more scattered, and they become more aggregated with increasing flow rates. It should be noted that, for the two gamma-mixed models, the parameter β is presented as its reciprocal value. This is due to the implementation of the gamma distribution in MATLAB, which is applied using the scale parameter, i.e. the reciprocal of rate parameter β. However, this does not have any side-effect on model estimations of interest.

In the two lognormal-mixed models, very similar results are found for μ and σ. There are no apparent differences found between the SPM and GQM models or among the three lanes. The parameters also seem more scattered at the low traffic flow level. It is worth noticing that the estimated values of μ in this study are very similar to what Branston (1976) studied on highway M4, where he reported μ of 0.39 for slow lane and 0.17 for fast lane. These are very close to estimated values of this study, for both lognormal-SPM and lognormal-GQM models.

For parameter ϕ, which is the proportion of followers, the estimations seem all similar among the four mixed models. With the increase of the traffic flow level, quite clear trends can be noticed in that the values of ϕ are also increasing. This is reasonable, since when the traffic become busier, more vehicles would be staying in queues as followers, so the proportion ϕ will increase. Among the three lanes, the estimations of ϕ on lane 3 are clearly larger than the estimations on other two lanes, which indicate that more vehicles in lane 3 are followers. This phenomenon is probably caused by the lane changing behaviours, as drivers are mostly looking for a gap to move into lane 3. In fact, the lower values of arrival rates λ support the lane changing behaviours too; since with lower arrivals, there should be lower traffic flows. However, those samples are all with high flow levels, which indicate that lane changing is a significant contribution to the traffic as well as the arrivals. The ϕ values of lane 2 is also larger comparing to lane 1, but this is not as obvious compared with the difference with lane 3.

One should notice that some of the estimations of ϕ are very close to 0, especially in lane 1 with flow levels less than 1300 veh/hr. Even though these estimates are well accepted by the GoF tests, they are clearly problematic. Some additional investigations were made such as varying the initial values, however, the problem remains. After carefully checking, it can be confirmed that the estimation method consistently gives the minimum values of the object function for MLE method. Hence, it is probably caused by the flexibility of the 4-parameter models, and the reasonable parameters may not be the best fit or the optimized estimations. Experiments have been done by fixing one or more parameters, and different estimation results were achieved; interestingly, most of the experimental fittings are also well accepted by GoF tests, or visual checking via the PDF graphs. Viewing the previous studies, Luttinen (1996) have commented that the mixed models might be too flexible, and he has suggested the use of a more powerful GoF test. However, a more powerful GoF test would not change or improve the result of parameter estimation, it only fails them. Ideally, giving more reasonable constraints on parameters may be an option to reduce the flexibility of models and make their estimation more meaningful. This requires better understanding of how the parameters relate to the traffic environment or conditions. The question of how to find better estimations to provide both good fittings and sensible interpretations to the headway data still remain open.

4.2.6 The Followers Headway

With respect to the construction of mixed models, the followers headway is a key element. Good followers headway models not only give better and more reasonable fits to empirical headway data, they also assist the analysis of vehicle following behaviours or on the explanation of some traffic phenomena using the parameters of the model.

Many researchers have previously studied followers headway models. Both Luttinen (1996) and Ha et al. (2012) have reviewed this aspect. More specifically, Wasielewski (1979) and Branston (1976, 1979) did analysis on this topic. Wasielewski used numerical integration to compute a non-parametric SPM model, and he found good fitting results on followers headway directly calculated from the empirical headway data. From the data he used, he concluded that the distribution of followers headway can be considered independently of the traffic flow level. He provided a result with a mean of 1.32s and a standard deviation of 0.52.

Branston used analytical forms of functions and tested them on both SPM and GQM models. He found similar GoF result for these two types of mixed models with the same analytical function. He also suggested that the parameters of followers headway distribution can be held constant over traffic flows without greatly changing the fit to the data. He pointed out that the constant parameters of followers headway distribution may overcome difficulties of parameter estimation at very low traffic flows.

TABLE 4.4: Mean Followers Headway (s)

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Gamma-						
SPM	1.33	0.26	1.15	0.22	1.18	0.11
GQM	1.32	0.25	1.16	0.21	1.18	0.12
Lognormal-						
SPM	1.56	0.81	1.38	0.31	1.26	0.14
GQM	1.61	0.47	1.43	0.32	1.25	0.14

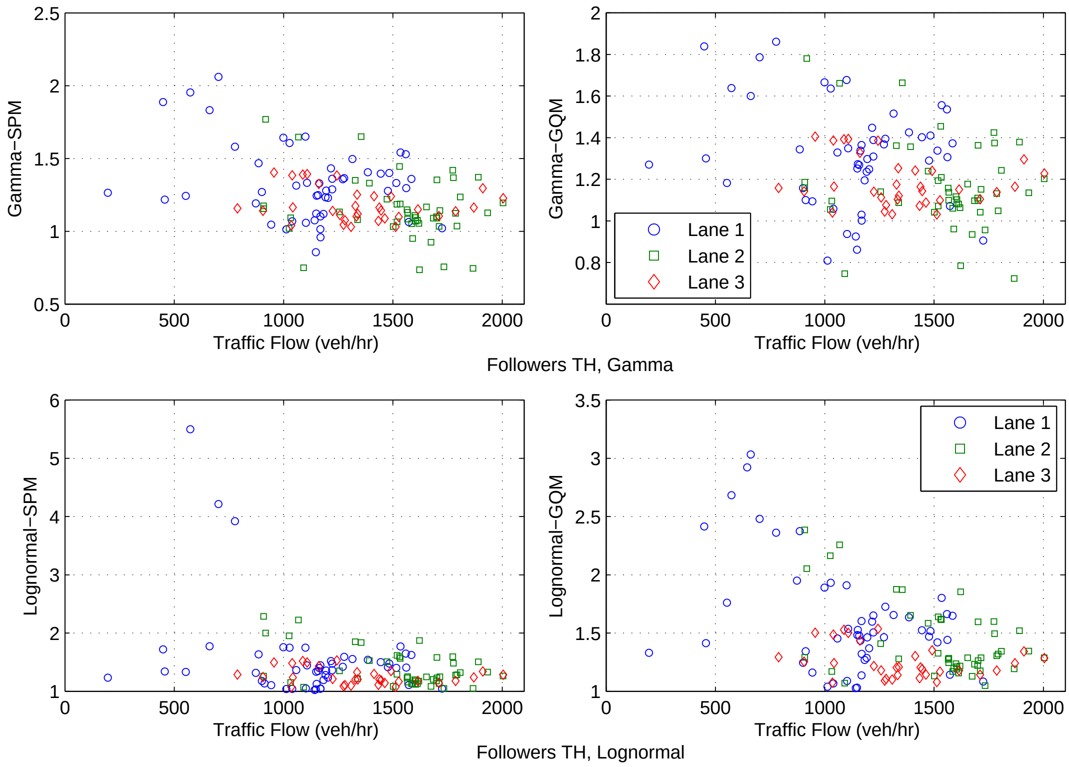

FIGURE 4.9: Mean followers headway (s) against traffic flows. The y-axis show the mean followers headway of each specified model.

In the present study, as shown in Figure 4.9, the mean followers headways in the gamma-mixed model are relatively consistent, with the range of 0.7s to 2.5s, where most of samples are with mean headway between 1s to 1.8s. The mean and standard deviation of each model are given in Table 4.4. The mean values of the lognormal-mixed model are similar to the result (mean of 1.32s and std of 0.52) of Wasielewski (1979), and to the result (1.62s for fast lane and 1.29s for slow lane) of M4 lane in the study of Branston (1976). The gamma-mixed models give slightly lower but close result compared to Wasielewski's result. Branston did not provide any result for gamma-mixed models.

The samples with higher mean values are mostly samples of lane 1 with flow rate less than 1000 veh/h. The highest bound is mostly less than 3.5s for the low flows. When the traffic becomes busier, the followers headway become tightly grouped with a highest bound less than 2s. This is similar to the discussion of Wasielewski (1979), who used 4s as the threshold to distinguish two vehicle groups. He suggested with 2.5s for flow less than 1400 veh/h and 3.5s for higher traffic flows, for the purpose of measuring empirical λ (the arrival rates).

Inspired by Wasielewski and Branston's work, similar investigations were carried out in this study in the following sections, to examine the effect of pre-determining some parameters of models as simplifications of the estimation process.

4.3 The parameter λ - Vehicle Arrivals

The parameter λ in mixed models is more appropriate to describe the tails of empirical headway density function, as it meant to represent vehicle arrival rates for of non-followers, or the gaps between traffic queues.

Imagining the vehicle on an idealised road, where they can all be free driving with similar speed, and all vehicle lengths are zero. If all drivers do not need a safe distance or say they can stop immediately to avoid any sudden collisions, then the distance between two consecutive vehicles can be infinitely small. Under this assumption, the vehicle numbers within a period of time will agree with a Poisson distribution, no matter how busy the road is. Hence the vehicle arrivals or time intervals between two vehicles will be a *negative exponential* distribution with parameter λ, where:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{4.33}$$

The mean arrival time is:

$$E[X] = \frac{1}{\lambda}$$

where E is the expectation of the random variable X for vehicle arrivals.

The correspondent hourly traffic flow rate is:

$$q = 3600\,\lambda(\text{veh/hr})$$

However, in reality, there are many restrictions on actual road traffic, as the vehicles have a range of lengths and they travel at different velocities, and drivers would not drive too close to the preceding vehicles because of safety concerns, reaction times, etc. As a result, the vehicles have to queue on the road to maintain a suitable gap, and this will cause a delay compared with the ideal arrival time

x in the above idealised circumstances. In this case, the traffic flow q estimated using λ will be larger than the real traffic flow, with the difference depending on traffic conditions, such as, how busy the road is. The busier the traffic, the longer queues will be, and they will form more frequently. Consequently, more delays will be caused. Thus the differences between q and realistic flow rate becomes greater. This explains phenomenologically why the negative exponential distribution can fit well for headways with lower traffic flows.

For the above reasons, using the assumption of exponential λ in mixed models, and consider the queuing vehicles as followers is a sensible method to describe the road traffic data if the chosen followers headway model is sufficient and the traffic flow is in relatively steady state. This difference between flows and ideal flows would have a relationship with traffic delays, and it can be used to check the followers headway models or may even be helpful for parameter estimation. However, This idea is not studied further here, but it is certainly worthy of further investigation.

This section is concentrated on the parameter λ, since if vehicles were successfully separated into follower and non-followers groups, the above discussion remains valid if only considering non-followers headway data. That is, the non-follower arrivals should be a *negative exponential* distribution, with the same parameter λ. The relationship between λ and its corresponding flow rates are presented using three linear equations for road lanes 1 to 3, respectively. Then these equations are used to predetermine the parameter λ in mixed models.

4.3.1 Previous studies of parameter λ

Previous studies have focused on the estimation of λ, or examining the exponential tail hypothesis. One of the main concepts is to find a headway threshold T, where all vehicles are non-followers if their headway is larger than T.

Miller (1961) studied a two-lane road in Sweden, he tested the threshold with 4, 5 and 6 seconds, and found that the T = 6s would be ideal. Luttinen (1996) tested data collected from so called 'low speed road' (50-70 km/h) and 'high speed road' (80-100 km/h) in Finland. He attempted to find T by testing where the exponential hypothesis holds, using GoF methods. He reported a T of 8 seconds for both low and high speed roads.

Wasielewski (1979) proposed a different method from Miller and Luttinen, and provided detailed algorithm in his paper. He iteratively checked the intervals between $T - 0.5$ s and T to search for significant deviation from the exponential determined by the headways with $t > T$. As mentioned in previous section, he found a $T = 2.5$s for flow $q < 1400$ veh/h and $T = 3.5$s for $q > 1450$ veh/h.

Other methods were studied using the speed difference to find threshold T (Branston, 1979; Bureau of Public Roads, 1950), which are not discussed here. Luttinen (1996) provided a comprehensive review.

4.3.2 Method to Predetermine parameter λ

The main barrier to estimating a threshold T using the headway data, is that the data at tail end is often very sparse especially with higher traffic flow. It causes fluctuations on estimation and it is hard to decide where the T should be located in those fluctuations. Also, if considering the road situation, the threshold T deviates from one data sample to another. To find a T for all data samples would be too idealistic. Such methods would be useful and necessary for some applications as other studies have addressed. However, finding a stable λ is the main focus of this study, so a different method are chosen here. Using the equation:

$$\lambda = \sum\nolimits_i (t_i - T)/n, \quad t_i > T \tag{4.34}$$

where the sum runs only when the headways t_i are greater than T and n is the total number of these headways.

The method is to iteratively calculate λ at T values from 1 to the maximum of t_is with an interval of 0.1 seconds. Then running averages of λ are calculated with an interval of 1 second. This will smooth the undesired fluctuations caused by the insufficient size of the data samples. The final step is to compare those running average levels and find a period of time where the mean λ does not significantly deviate. The result shows that a running average between 2.0 and 3.5 s can give a stable λ.

Using the above method, λ values for all samples are calculated and grouped by lane 1-3. Then linear regression for each group of lambdas, 3 equations are achieved, as shown in Table 4.5.

Table 4.5: Estimation of λ as function of traffic flow q

Parameter	Lane 1	Lane 2	Lane 3
a	1.509	1.607	0.736
b	-0.045	-0.102	0.012

a, b are parameters of function $\lambda(q)$, where $\lambda(q) = a * q + b$, q is sample flow, measured in veh/second.

4.3.3 Parameter Estimations with given λ

Based on the equations in Table 4.5, λ can be determined using the sample traffic flow. Both gamma-SPM and gamma-GQM are estimated to see how predetermined λ can affect headway data fitting. The lognormal-mixed models are not tested from this point onward, due to their high computation times.

The estimation method is the same for the rest of the three parameters, using the numerical MLE method.

Result and discussion

After parameter estimation, the result of the GoF test on both KS and AD test are presented in Figure 4.10, and summarized in Table 4.6. Results of both gamma-SPM and gamma-GQM show the estimation of these two models are very well accepted. Two more samples in lane 1 did not pass the KS test, compared to the method in the previous section (see Table 4.2). This indicates that pre-determining λ before estimation does not change fitting result much.

TABLE 4.6: GoF results with pre-determined λ on gamma-SPM and gamma-GQM models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
Gamma-						
SPM	49(3)	51(1)	47(0)	47(0)	33(0)	33(0)
GQM	49(3)	51(1)	47(0)	47(0)	33(0)	33(0)

The numbers with or without brackets show the numbers of successes or failures of GoF test.

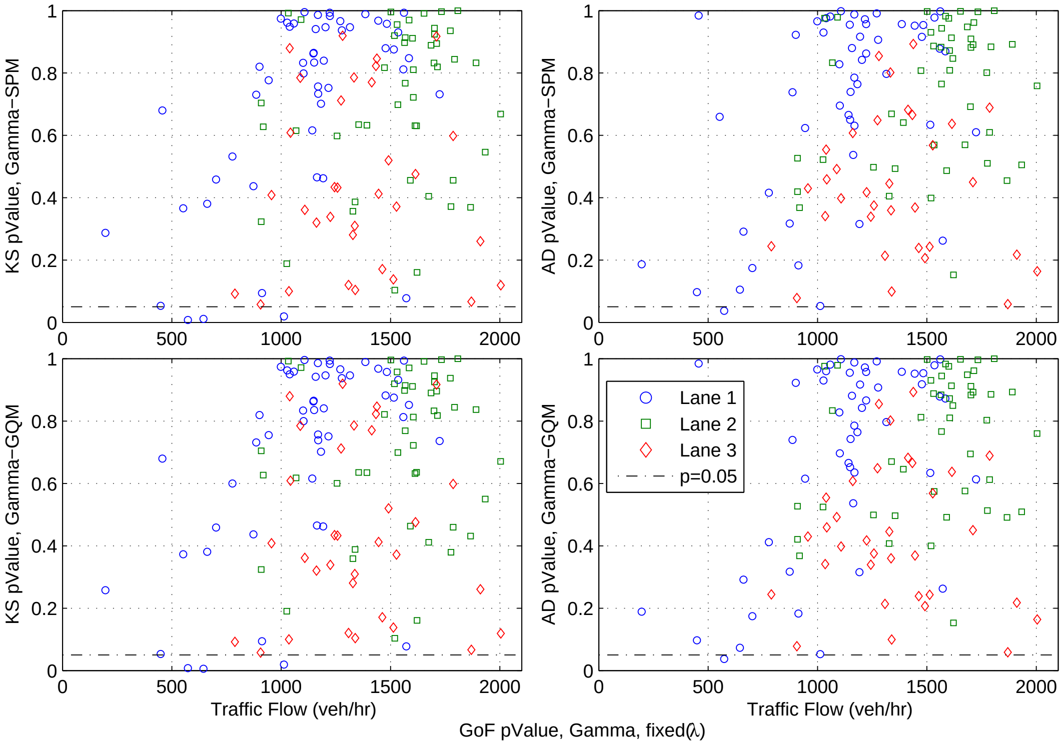

FIGURE 4.10: GoF results (p value) with pre-determined λ on gamma-SPM and gamma-GQM models

The estimation time, as show in Table 4.7, is effectively reduced about 40% on both models. This is not surprising as the parameter λ is required to fit in the estimation stage.

The estimated parameters with pre-determined λ are similar to the results in section 4.2.5. As shown in Table 4.8, the mean and standard deviation of ϕ, α and β have similar ranges compared to Table 4.3. However, in lane 1, there are

Table 4.7: Estimation Time with pre-determined λ on gamma-SPM and gamma-GQM models

	Fitting Time (s)		KS Test (ms)		AD Test (ms)	
	Mean	Std	Mean	Std	Mean	Std
Gamma-						
SPM	0.60	0.27	5.15	2.93	4.00	1.76
GQM	0.56	0.28	4.57	2.73	3.45	1.47

larger standard deviations of more than 1 of β, comparing to the previous of 0.1; this is because that one sample (No. 189) had extreme estimated values of these 3 parameters. With the very low flow rate (195 veh/hr), the estimation of 0.7 for ϕ, seems unreasonably high. Although it can pass GoF test, the estimation of this sample is problematic. This is probably caused by the inaccuracy of equations on λ, which is not good enough to give a suitable value at very low level of traffic flows. If the sample (No. 189) is treated as an outlier, the result (see the values in brackets) in lane 1 become become similar range as in Table 4.3.

Table 4.8: Estimated Parameters with pre-determined λ on gamma-SPM and gamma-GQM models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Gamma-SPM						
ϕ	0.19	0.16	0.29	0.11	0.57	0.08
λ	0.43	0.13	0.58	0.12	0.29	0.06
α	10.30(10.48)	4.27(4.12)	9.98	1.88	9.08	1.83
$1/\beta$	0.30(0.15)	1.05(0.06)	0.12	0.04	0.14	0.04
Gamma-GQM						
ϕ	0.23	0.17	0.33	0.12	0.57	0.08
λ	0.43	0.13	0.58	0.12	0.29	0.06
α	10.52(10.71)	4.23(4.07)	10.02	1.87	9.09	1.83
$1/\beta$	0.30(0.14)	1.10(0.06)	0.12	0.04	0.14	0.04

For samples in lane 1, as shown in Figure 4.11 and 4.12, there are still a number of estimations of ϕ that are close to 0, mentioned earlier. These estimations are

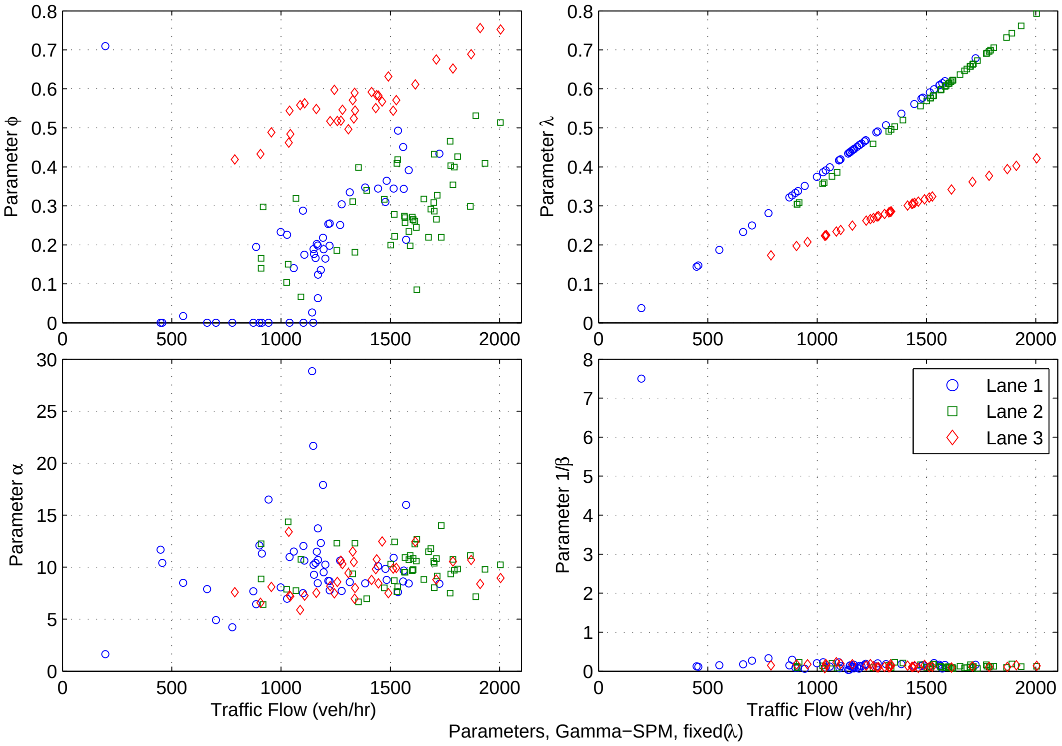

FIGURE 4.11: Estimated parameters with pre-determined λ on gamma-SPM models

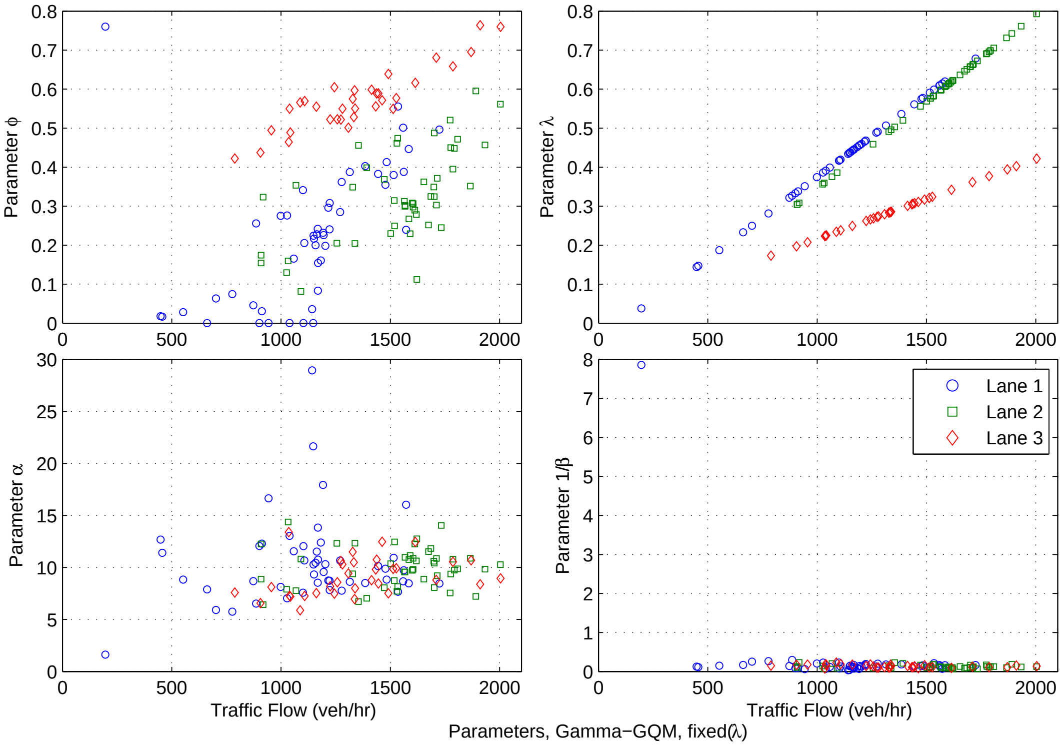

FIGURE 4.12: Estimated parameters with pre-determined λ on gamma-GQM models

still unreasonable, since there must be some proportion of followers travelling in queues at corresponding traffic flow levels.

The mean followers headway have not varied much compared with the previous section either. Again, one sample (No. 189) had a very high mean headway of 12 seconds, which was caused by the problematic estimation happened in lane 1. This also was caused by much larger values of mean and standard deviation across all samples of lane 1, as shown in Table 4.9. Taking away this outlier, Table 4.4 and 4.9 are similar. In lane 2, the standard deviation reduced, which may indicate that the mean followers headway becomes more aggregated with the pre-determined λ.

TABLE 4.9: Mean Followers Headway (s) with pre-determined λ

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Gamma-						
SPM	1.55(1.32)	1.58(0.21)	1.17	0.16	1.18	0.11
GQM	1.58(1.34)	1.65(0.22)	1.18	0.16	1.18	0.11

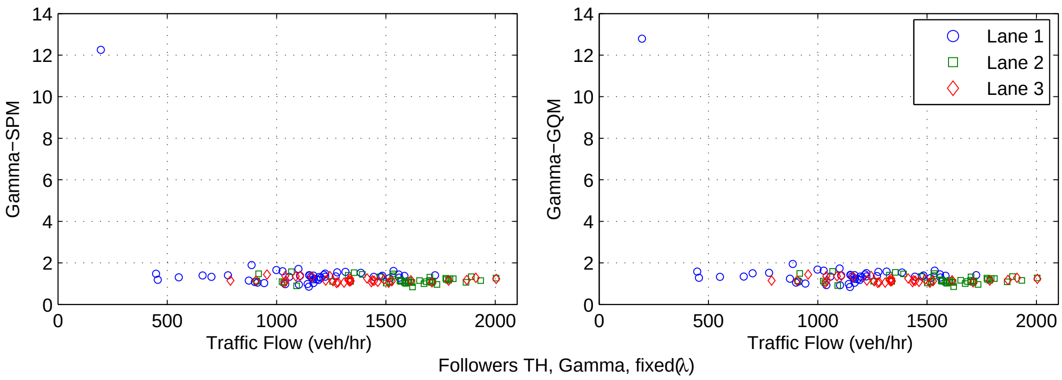

FIGURE 4.13: Mean followers headway with pre-determined λ on gamma-Mixed Models. The y-axis show the mean followers headway of each specified model.

Overall, with pre-determined parameter λ, the speeds in fitting stage are improved, without much variation in the performance of model fitting to the empirical headway data. This positive result shows the reasonability of using arrival rates of λ to improve the estimation of gamma-mixed models.

The results shown above are very promising, as firstly this method helps in fitting mixed models to headway data without affecting their fitting performance. Secondly, these results support well as additional evidence on the exponential tail hypothesis discussed earlier. Thirdly, as λ is estimated using sample flow rates, rather than model fitting, it can be used for mixed and presumably combined models with any kind of followers headway distribution. Moreover, the different linear expressions of λ among varies lane reveal that this parameter is highly dependent on traffic environment, and this can be further investigated with more headway data available, especially in various locations and road figures. Such investigations would potentially identify external factors that affect the changes of λ, and this will be valuable for the analysis of traffic flows.

4.4 Attempts on More Simplified Estimation

4.4.1 Estimations on Fixed Followers Headway Models

Researchers previously attempted to use fixed followers headway models to simplify the estimation process of SPM and GQM mixed models. Wasielewski (1979) reported good results on his non-parametric followers headway SPM model, which he used an independent followers headway model evaluated directly from headway data. He found that all samples have passed the GoF test. Branston (1976) tested the lognormal-GQM model as a parametric model. He took average values categorized by lanes and loop locations, and used those values as the parameters of a *lognormal* followers headway model. He reported that all the estimation results are well accepted by the GoF test and concluded that a constant followers distribution can be an acceptable simplification in fitting the lognormal-GQM to headway data.

The above studies are good indications that using a fixed followers distribution can be a reasonable alternative method on simplifying mixed models. Inspired by this, some similar attempts are investigated in this thesis.

The first apparent step is to repeat Branston's experiment using gamma-mixed and lognormal-mixed models with headway data used in the present study. As discussed in section 4.2.5, the mean estimated values of μ and σ are very close to what have been reported by Branston, particularly on data for the M4 motorway. Following his procedure, parameters α and β in the *gamma* distribution, as well as the μ and σ in the *lognormal* distribution are fixed for each of the lanes 1-3. Model fittings are then implemented to fit headway samples.

The results of the GoF test for these simplifications are listed in Table 4.10. Generally, the results on lane 1 and 2 are acceptable. However, there is a big performance drop in lane 3, where only about half of the samples have passed GoF, which is not ideal. By the time Branston collected his data from a overhead section of the M4 motorway, which had only two lanes on each direction, he did not have data of lane

3. From the test results, the traffic in lane 3 may be more complex to use a model which is simplified by only using a fixed followers headway distribution. Due to the low performance on lane 3, the detailed estimation results are not presented in this thesis. In addition, the computational costs of using lognormal-GQM are still unbearable. Also, the computational costs on the GoF test for both lognormal-mixed models are high, due to the numerical integration in computing CDF of lognormal-mixed models. This may sound a bit negative with the downgraded fitting in lane 3. However, the similar results of lanes 1 and 2 compared with Branston's study, suggests that the fixed followers headway model is useful with traffic that is not heavily interrupted by lane changing vehicles, which is probably the cause of the degraded fitting performance.

TABLE 4.10: Summary of GoF results using fixed followers headway models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
Gamma-						
SPM	49(3)	50(2)	47(0)	47(0)	16(17)	16(17)
GQM	50(2)	50(2)	47(0)	47(0)	16(17)	16(17)
Lognormal-						
SPM	50(2)	50(2)	46(1)	46(1)	18(15)	20(13)
GQM	50(2)	52(0)	46(1)	46(1)	18(15)	17(16)

4.4.2 Alternative Simplification Methods

Considering that using the pre-determined parameter λ is a good method of simplification, the combinations of λ with one of the parameters in followers headway distribution may be alternative methods to simplify the estimation process. This study tested a few ways of combining parameters for the two gamma-mixed models.

After some trials, it was found that using the combination of λ and α gives promising results. This method was again using equations in Table 4.5 to pre-determine λ, and also using the average of α estimated earlier, with a value of 8.3, for all the three lanes. Both gamma-SPM and gamma-GQM were tested using this method.

TABLE 4.11: Summary of GoF results with pre-determined λ and α

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
Gamma-						
SPM	49(3)	50(2)	47(0)	47(0)	32(1)	31(2)
GQM	48(4)	50(2)	47(0)	47(0)	32(1)	31(2)

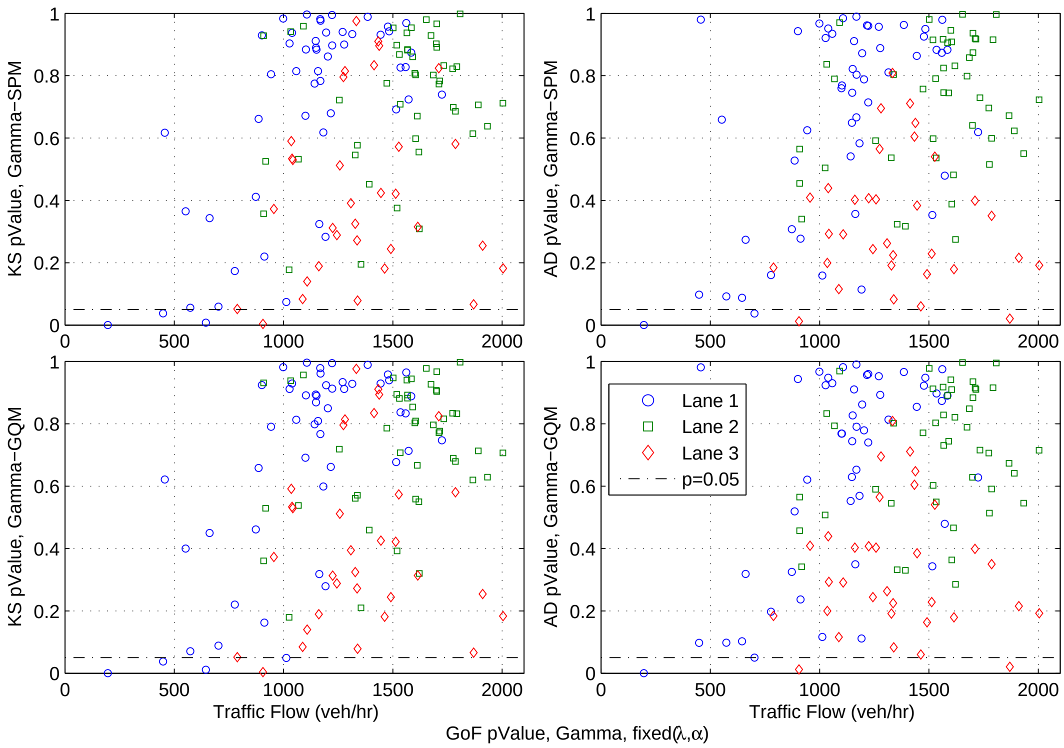

FIGURE 4.14: GoF results with pre-determined λ and α on gamma-mixed Models

Table 4.11 and Figure 4.14 show a summary of the results of the GoF tests. The estimations of all samples have passed GoF tests in lane 2. In lanes 1 and 3, most of the samples are well accepted, but the results are slightly worse compared with Table 4.2 and 4.6, with 1-2 more samples rejected by the GoF tests. Most of the rejected samples have occurred on a relatively low flow level of below 1000 veh/hour.

With an additional parameter reduced in estimation process, the fitting time (Table 4.12) again reduced by more than half, with the mean estimation time close to 0.25 seconds.

TABLE 4.12: Comparison of estimation time(s) on gamma-SPM, gamma-GQM, with pre-determined λ and α

	Fitting Time(s)		KS Test (ms)		AD Test (ms)	
	Mean	Std	Mean	Std	Mean	Std
Gamma-						
SPM	0.26	0.08	4.91	2.05	4.04	1.75
GQM	0.24	0.07	4.27	1.94	3.35	1.37

Table 4.13 and Figure 4.15 and 4.16 show the result of estimation parameters with this alternative method. It is worth noticing that the parameter β is much tighter to its mean values compared with the previous results in Table 4.15 and Table 4.16. The outlier sample (No. 189) is rejected GoF test this time.

The estimation of ϕ is improved by this simplification, as some of above-mentioned problematic values deviate from zero. Also, for each lane, the figures show clear increasing trends of the proportion factor ϕ with increased traffic flows. This gives a better sense of the nature of traffic on the real road, considering the proportion of queues that depend on traffic flows. Wasielewski (1979) had his data collected

TABLE 4.13: Estimated Parameters with pre-determined λ and α on gamma-SPM and gamma-GQM models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Gamma-SPM						
ϕ	0.21	0.13	0.32	0.10	0.58	0.08
λ	0.44	0.12	0.58	0.12	0.29	0.06
α	8.30	0.00	8.30	0.00	8.30	0.00
$1/\beta$	0.17	0.02	0.15	0.01	0.14	0.01
Gamma-GQM						
ϕ	0.27	0.14	0.37	0.10	0.58	0.08
λ	0.44	0.12	0.58	0.12	0.29	0.06
α	8.30	0.00	8.30	0.00	8.30	0.00
$1/\beta$	0.18	0.01	0.15	0.01	0.14	0.01

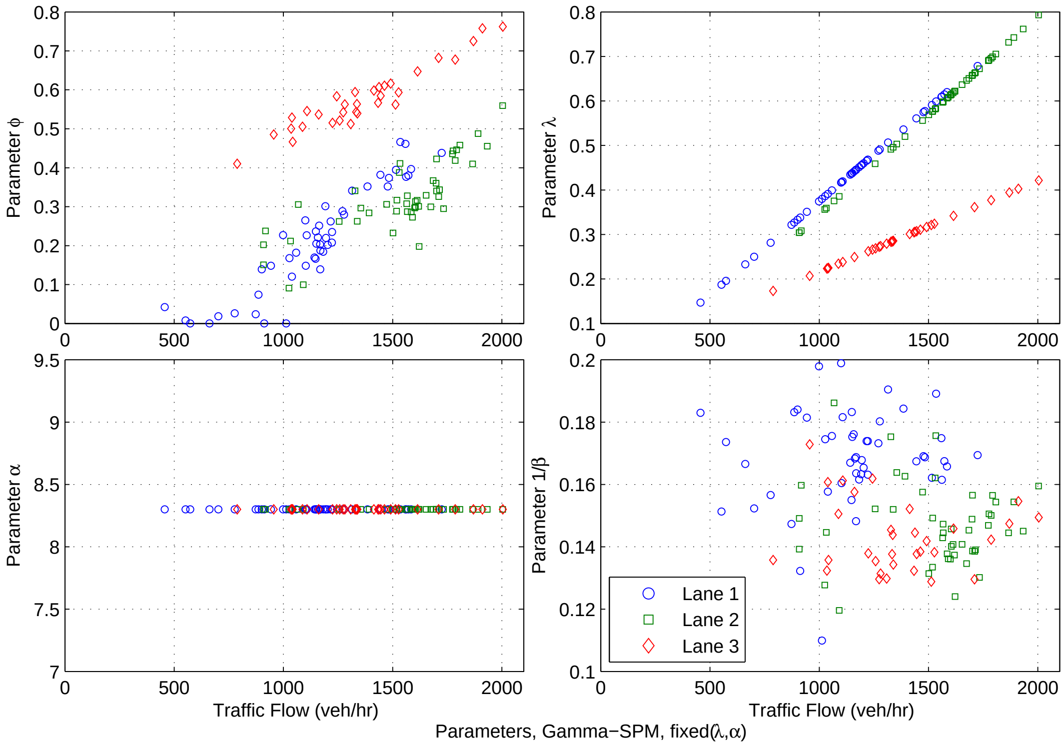

FIGURE 4.15: Estimated parameters with pre-determined λ and α on gamma-SPM models

from a Detroit freeway, and reported a similar trend without grouping data by lanes. In contrast, a clear distinction across lanes is shown in this study. Lanes 1 and 2 have a similar trend of ϕ, and the difference does not seem significant. Estimated values in lane 3 are much higher for corresponding traffic flows. This

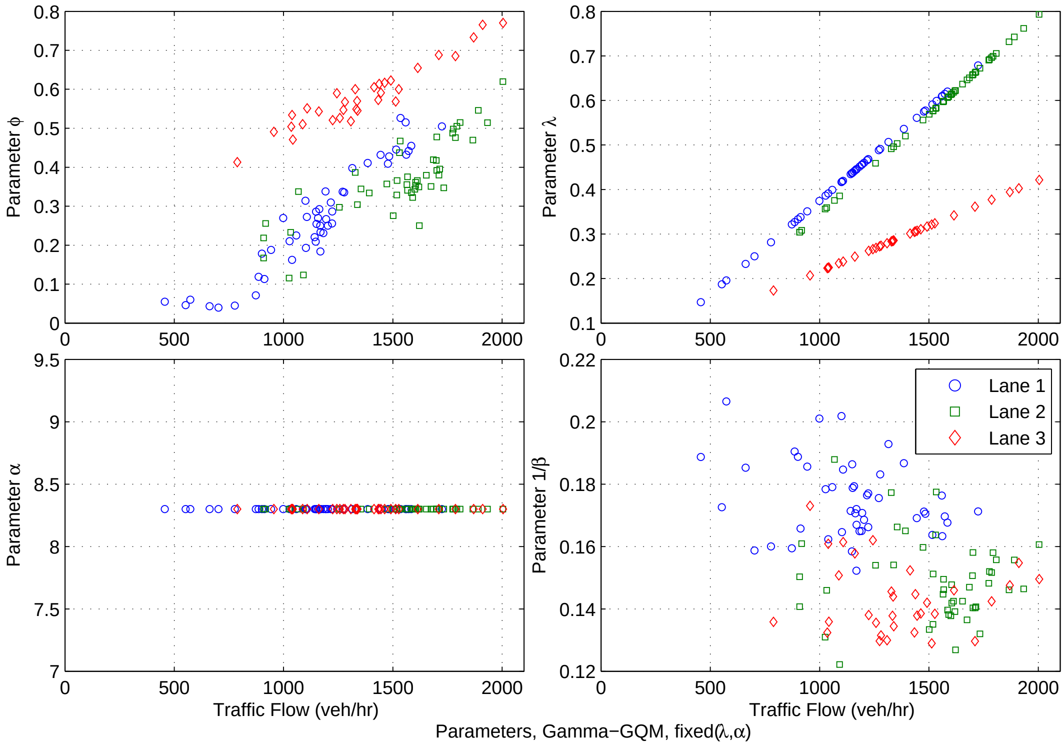

FIGURE 4.16: Estimated parameters with pre-determined λ and α on gamma-GQM models

phenomenon may suggest that more vehicles in lane 3 are gathered in queues other than driving individually. This gives a similar hint compared with the much lower λ values in lane 3, as discussed in section 4.3, a good proportion of vehicles may have changed their lanes and they will find a suitable gap to join lane 3, and this will form more queues in traffic.

The mean followers headways are more aggregated to their means across all samples. In relatively lower traffic flows, the mean values seem scattered, and they become more condensed with increased traffic flows. In Table 4.14 and Figure 4.17, differences between lanes can be noticed, where the values in lane 1 are larger and more scattered, while lane 2 and 3 are similar, especially when traffic flows are more than 1500 veh/hr. It may be explained by that, when the road is busy, and queues become more condensed, drivers may tend to have similar following behaviours.

TABLE 4.14: Mean followers Headway (s) with pre-determined λ and α

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Gamma-						
SPM	1.40	0.13	1.22	0.11	1.19	0.09
GQM	1.45	0.10	1.23	0.11	1.19	0.09

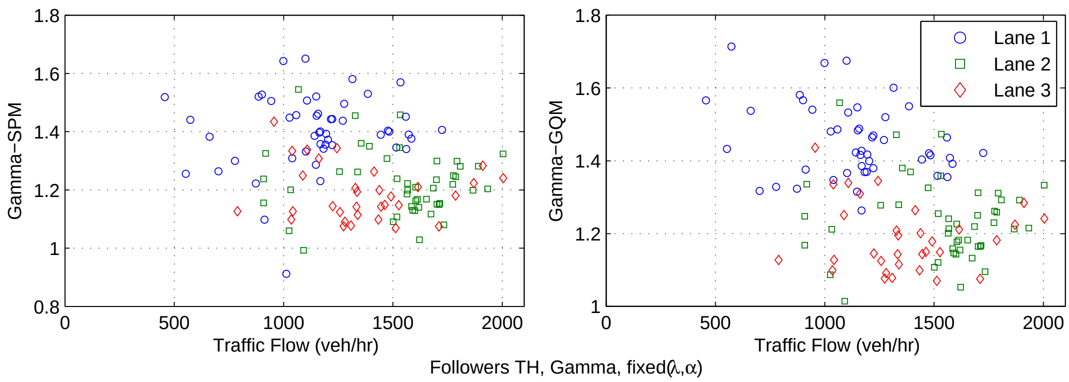

FIGURE 4.17: Mean followers headway (s) with pre-determined λ and α on gamma-Mixed Models. The y-axis show the mean followers headway of each specified model.

Compared to the method of using fixed followers headway model in mixed models, using a combination of pre-determined λ and fixed α in gamma-mixed models has improved fitting results especially in lane 3. It has also improved the estimation of ϕ, as some of previously problematic estimations seem more reasonable. With the fixed α, the mean followers headways become more consistent with smaller variations. This is similar to values (within range 1.2-1.4 seconds) reported by Wasielewski (1979), although the fixed followers headway model did not help to fit data well in some traffic conditions (lane 3 in this case). Clearly, the suggested method of this section is valid and useful when using mixed models in analysis of headway data.

4.4.3 Considering Parameter ϕ for Simplification

The result of the previous section shows good increasing trends of ϕ with increased flows. If ϕ can be found to have good relationship with flow rates q, then it will help to accurately estimate proportion of queue in a specific traffic condition. By using quadratic methods, three quadratic equations are obtained to describe the trend of ϕ as the function of flow rate q, as shown in Table 4.15.

Based on method used in the previous section, an additional parameter ϕ is also added to the combination, i.e. to pre-determining ϕ, λ and α, as a final attempt to simplify the estimation of mixed models.

TABLE 4.15: Estimation of ϕ as function of traffic flow q

	Gamma-SPM			Gamma-SPM		
	Lane 1	Lane 2	Lane 3	Lane 1	Lane 2	Lane 3
a	3.28	3.45	0.93	3.01	3.09	0.93
b	-0.52	-1.73	0.25	-0.24	-1.30	0.25
c	0.03	0.41	0.35	0.02	0.34	0.35

a, b and c are parameters of function $\phi(q)$, where $\phi(q) = a * q^2 + b * q + c$, and q (veh/second) is sample traffic flow.

A similar estimation used with ϕ pre-determined by equations in Table 4.15, as well as λ and α, which are discussed in the previous section. The fitting results are examined with the GoF test.

TABLE 4.16: GoF results with pre-determined ϕ, λ and α on gamma-SPM and gamma-GQM models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
Gamma-						
SPM	48(4)	49(3)	44(3)	45(2)	27(6)	29(4)
GQM	48(4)	49(3)	43(4)	45(2)	27(6)	28(5)

The numbers with or without brackets show the numbers of successes or failures of GoF test.

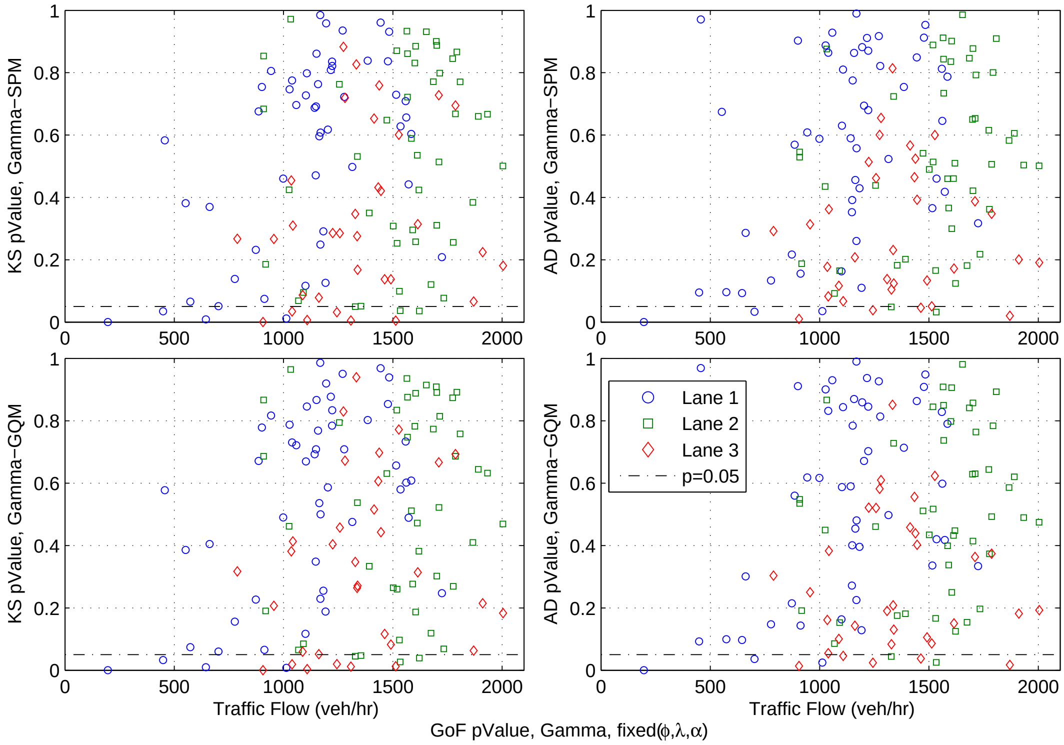

FIGURE 4.18: GoF results (p value) with pre-determined ϕ, λ and α on gamma-SPM and gamma-GQM models

The results show reasonable acceptances of GoF tests in Table 4.16 and Figure 4.18, and there are a few more rejected samples compared to previous sections. In lane 1, the result are slightly worse compared with both gamma-mixed models. More samples are rejected by the GoF test in lane 2 and 3. Particularly in lane 3, where 6 out of 33 samples did not pass the KS test. With more restricted parameters, gamma-mixed models become less flexible and less capable of fitting headway data.

The results of the estimated parameter β, in Table 4.17, have very similar range compared with the result in Table 4.13, that is, method of pre-determined ϕ has not affected much the estimated parameters (only β in this case). This also applies to the mean followers headways, which also are not changed by this method.

TABLE 4.17: Estimated Parameters with pre-determined ϕ, λ and α on gamma-SPM and gamma-GQM models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Gamma-SPM						
ϕ	0.22	0.12	0.32	0.08	0.59	0.08
λ	0.44	0.12	0.58	0.13	0.30	0.06
α	8.30	0.00	8.30	0.00	8.30	0.00
$1/\beta$	0.17	0.01	0.15	0.01	0.14	0.01
Gamma-GQM						
ϕ	0.27	0.13	0.37	0.09	0.59	0.08
λ	0.44	0.12	0.59	0.13	0.30	0.06
α	8.30	0.00	8.30	0.00	8.30	0.00
$1/\beta$	0.18	0.01	0.15	0.01	0.14	0.01

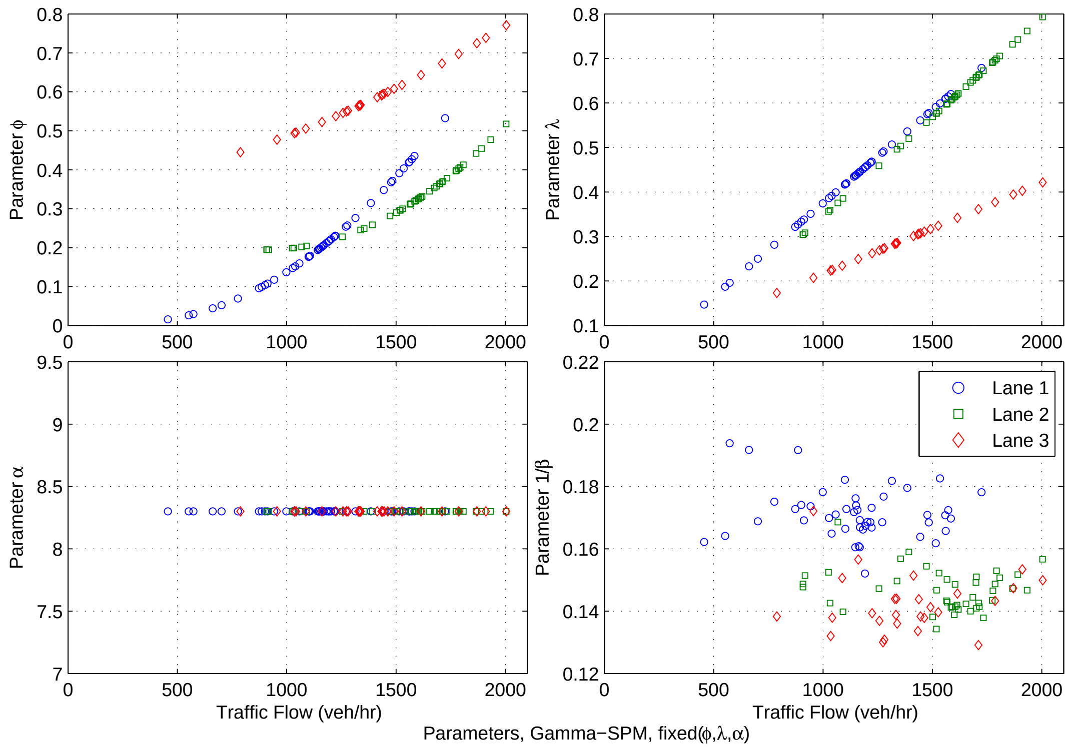

FIGURE 4.19: Estimated parameters with pre-determined ϕ, λ and α on gamma-SPM models

Although pre-determining ϕ does not give excellent fitting as previous methods do, the overall good acceptance still suggests that parameter ϕ can potentially be used to simplify the mixed models with more careful analysis and probably with more data from different sites, to achieve excellent fitting results. If this can

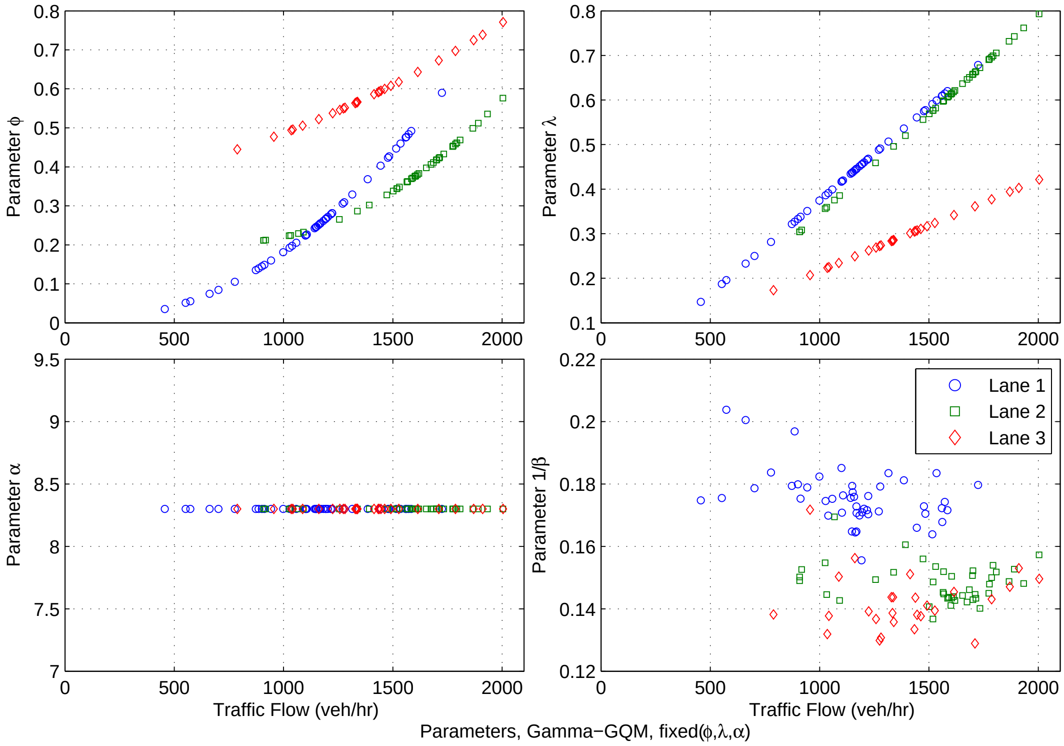

FIGURE 4.20: Estimated parameters with pre-determined ϕ, λ and α on gamma-GQM models

TABLE 4.18: Mean Followers Headway (s) with pre-determined ϕ, λ and α

Gamma-	Lane 1 Mean	Lane 1 Std	Lane 2 Mean	Lane 2 Std	Lane 3 Mean	Lane 3 Std
SPM	1.42	0.07	1.22	0.05	1.18	0.08
GQM	1.46	0.07	1.23	0.05	1.18	0.08

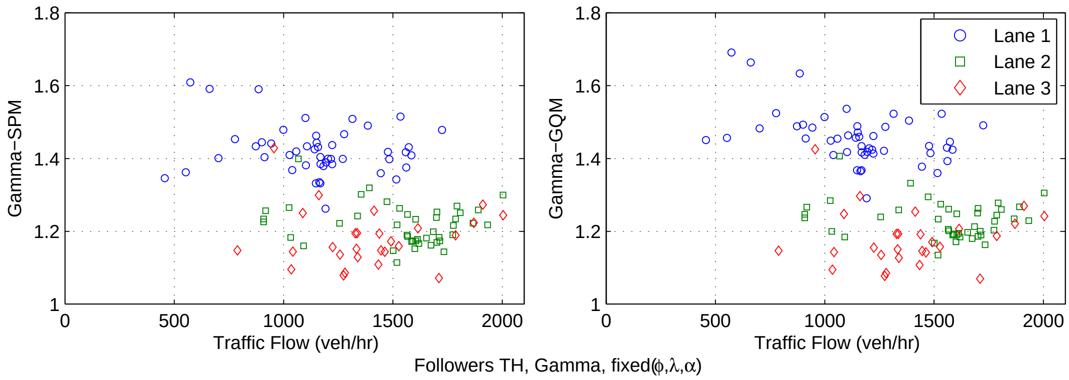

FIGURE 4.21: Mean Followers Headway (s) with pre-determined ϕ, λ and α on Gamma-Mixed Models. The y-axis show the mean followers headway of each specified model.

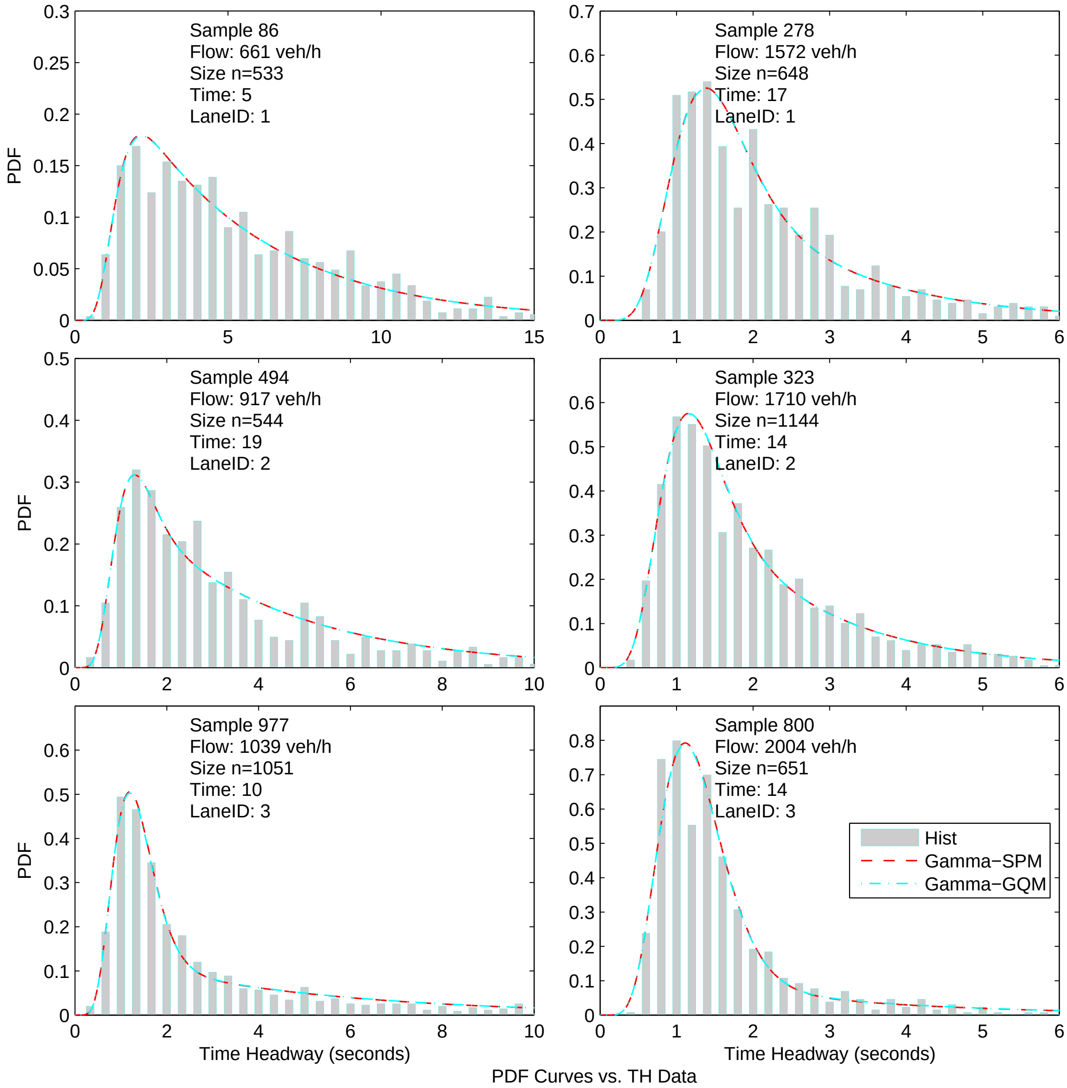

FIGURE 4.22: PDF curves with pre-determined ϕ, λ and α on gamma-SPM models

be generalized, it can be very useful for explaining the queue formulations and drivers' car following behaviours.

The main downside is that ϕ is sensitive to the mixed models selected, even when with same followers headway models, ϕ is still required to be estimated differently between SPM and GQM models. This will badly affect the generalization of this method, as this parameter is not model independent. Also, this model dependency of ϕ shows that the estimated proportion of queues are also highly dependent on the models chosen. This also makes estimated models unrealistic, as there is only one true proportion of queues in reality. Further investigations with more data may find more about this parameter, and it may potentially be useful in not only understanding more about queues, but also in validating the mixed models being used.

4.5 Review and Discussion

The mixed distribution models are widely used in the analysis of time-headway data. There are major two types of mixed models, i.e. SPM and GQM. This chapter presents a brief discussion of the use of these two models.

Gamma- and lognormal-mixed models are tested with data samples used in this study, and very good fitting results have been achieved in the terms of the GoF tests. The results are similar to findings of previous studies. Importantly, the lognormal-SPM model has been, for the first time, implemented by using an approximation method for Laplace transform of *lognormal* distribution. In both followers headway models, i.e. the *gamma* and *lognormal*, very similar results of parameter estimation are achieved between the SPM and GQM, which indicate the quantitative similarities between these two type of mixed models. The computation costs on estimation and GoF tests are much higher for lognormal-mixed models, due to the unavoidable numerical integrations to solve PDF and CDF, especially in the GQM case. The speed on estimating lognormal-SPM seems better than lognormal-GQM, but still much worse comparing to estimations of gamma-mixed models.

Considering estimated parameters, a good relationship is found between the arrival rates λ and traffic flow levels, and this helps further in the simplification of fitting mixed models. All the estimated parameters of followers headway models, i.e. α and β of the *gamma* distribution, as well as μ and σ of the *lognormal* distribution, seem to be independent of traffic flow. They are more scattered in the lower level of flows, but aggregated more densely when the flow level increases. They are also distinguishable among the lanes of 1-3. Some of the parameter ϕ, the fraction of followers group, are estimated very close to zero, especially in lane 1. These are unreasonable estimates, hence should be considered as problematic estimations, despite the good fitting results achieved.

The fixed followers distribution method is also tested for both the gamma- and lognormal-mixed models, as previous studies reported such with good indication. Using this method, the fitting results in lanes 1 and 2 are very well accepted by the GoF test, as expected. However, the traffic in lane 3 seems more complex, and only about half of the estimations are well fitted to headway data. Although, it is yet unclear of the causes for the downgraded fitting in lane 3, but the similar results found compared with previous studies in lane 1 and 2 supports very well that this method can be very useful under certain traffic conditions.

Some methods of simplification on estimation of gamma-mixed models are attempted. The positive test results show that using pre-determined λ, combined with a constant α can effectively improve the fitting process, without sacrificing their performance or great variation of parameters. The method of using the combination of λ and α shows some level of correction on the estimation of parameter ϕ, and hence reveals positive trends between ϕ and traffic flows, which supports the concept that parameter ϕ is presenting the proportion of followers group. This is further supported by using ϕ as an additional pre-determined parameter for the estimation.

In general, as four-parameter models, gamma-mixed models are overly flexible. This may lead some problematic estimations as many parameter combinations can be acceptable result, in terms of the GoF test. This problem may not be solved by giving better estimation algorithms or making a more powerful GoF test, as an optimum solution may not always be the most sensible solution. The attempts on simplifying the models show a good potential of solving this problem, by effectively restricting them using external conditions such as flow rates. Indeed, giving reasonable constraint on some of the better understood parameters would narrow the scope of solutions, hence force the estimation to gain a better result, presumably closer to the reality. To accomplish this target, more information and investigations are desirable for better understanding of the drivers' following behaviours.

Chapter 5

The Use of Response Time Models in Time-Headway Analysis

Summary

This chapter introduces two distribution models, the exponential modified Gaussian (EMG) and inverse Gaussian (IVG) distributions, which are often applied as response time (RT) models in cognitive psychology. Both models are explained and then fitted, as single models, using empirical headway samples. Furthermore, with the two models used as followers headway distributions, all SPM and GQM mixed models are implemented and tested. The GoF results show excellent fitting performance with the use of both EMG and IVG distributions. In the later sections, methods of simplifying parameter estimations are attempted and discussed.

5.1 Response Time Models

5.1.1 Introduction of the Response Time Models

In a steady-state car following situation, a time-headway can be seen as a sum of two components. The first component is the time needed for the preceding vehicle to pass a test point. This time component depends on its speed and length. The second time component is a time gap between the two vehicles, and this gap allows the following vehicle to drive with a safe distance so that it can react to any potential changes on road, such as noticing a slowing down of the preceding vehicle and then pushing down the brake pedal. This time gap can be seen as a the follower's comfort gap where it can follow the preceding vehicle safely based on the driver's own perception and judgement. Although there is lack of information on how this time gap is determined, researchers often use the reaction time as an important factor of it, in car-following theories (Forbes, 1963; Forbes and Simpson, 1968; Forbes et al., 1958; May, 1990).

The reaction time used in car-following is often known as perception-brake reaction time (Green, 2000) or brake reaction time(BRT) (Johansson and Rumar, 1971). This brake reaction time measures the time taken from drivers' perception, to their actions of braking after a decision has been made. There is intensive literature for measuring such a time as drivers braking performance (Lerner, 1993; Sivak et al., 1981; Summala, 2000; Young and Stanton, 2007). For example, Johansson and Rumar (1971) reported the estimated reaction time varied from 0.4 to 2.7 s, with a mean value of 1.0 s. In their experiment, subjects are instructed to press the brake pedal as soon as they hear a sound.

Roughly speaking, this brake reaction process is analogous to response time (RT) of taking a two-choice task in cognitive psychology, as both processes apply the procedures of perception, decision making and taking actions. Since the analysis of response time have been widely studied with a long history (Hohle, 1965; Luce, 1986; Matzke and Wagenmakers, 2009; Ratcliff and McKoon, 2008; Townsend and Ashby, 1983). It is assumed that some of the good models used in RT analysis can also be used in the analysis of vehicle time-headways. Although there is no theoretical reasoning to support such an assumption, the fact that *gamma* and *lognormal* distributions are also used as RT models (Baayen and Milin, 2010; Ratcliff and Murdock, 1976; Ulrich and Miller, 1993; Van Zandt, 2000) gives a kind of intuition that RT models might as well be an option to time-headway analysis.

Among many of the RT distribution models, the exponential modified Gaussian (EMG for short, which is often named as ex-Gaussian), inverse Gaussian (IVG for short, which is also know as Wald distribution) and shifted-IVG (or shifted-Wald) models are widely recognized as good performance models in RT analysis (Hockley, 1984; Hohle, 1965; Luce, 1986; Matzke and Wagenmakers, 2009; Van Zandt, 2000).

In the first part of this chapter, EMG, IVG and shifted-IVG models are fitted using the headway samples and their results are compared with *lognormal* models discussed in chapter 2. This attempt is to test whether these selected distributions are suitable for headway analysis. However, it is important to note such attempt does not indicate that headway processes are necessarily comparable to RT processes, which need much deeper studies that involve more theoretical and experimental investigations. Nevertheless, the successful fitting of the selected models would give research more useful options when describing and hence predicting headway data statistically.

5.1.2 Models Description

Similar to the implementations in Chapters 2 and 4, the model fitting methods are used the numerical MLE to estimate the parameters. Then models are tested using goodness-of-fit test to examine the fitting result as part of the fitting performance. To this extend, the PDF and CDF functions are used in the proposed framework.

The mean and variance properties are also important properties especially when models are used as the followers headway distribution models, as they are the estimations of the mean and variance of the followers headways, which are difficult to measure directly from the empirical data. Also, the mean and variance can be used as first- and second-order moments that will give a comparison between values of empirical measurement and model estimations.

The Laplace transform of models is very important when the models are used in the SPM mixed models, so they are also listed as properties considered in this study. There are more statistical properties of these selected models. Readers can find more details via the references given.

Exponential Modified Gaussian(EMG) distribution

For RT data, the most common parametric estimate of the PDF is the EMG model, the convolution of a normal and an exponential random variable (Van Zandt, 2000).

The properties of EMG distributions refer to Haney (2011). Haney studied some practical applications and properties of EMG model, and he also gives the details of derivation of PDF and CDF of it in his thesis.

PDF:

$$g(t;\mu,\sigma,\tau) = \frac{\tau}{2} e^{\tau\left(\mu - t + \frac{\tau\sigma^2}{2}\right)} \operatorname{erfc}\left(\frac{\mu - t + \tau\sigma^2}{\sqrt{2}\sigma}\right) \qquad \mu > 0, \sigma > 0, \tau > 0 \tag{5.1}$$

where $\operatorname{erfc}(x)$ is complementary error function with

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \mathrm{d}t$$

For the convenience of using this PDF in later mixed models, the function is also notated as $g_e(t;\mu,\sigma,\tau)$, which indicate that is the PDF of EMG as followers headway distribution.

CDF:

$$G(t;\mu,\sigma,\tau) = \frac{1}{2}\left(\operatorname{erfc}\left(\frac{(\mu - t)}{\sqrt{2}\sigma}\right) - e^{\tau\left(\mu - t + \frac{\tau\sigma^2}{2}\right)} \operatorname{erfc}\left(\frac{\mu - t + \tau\sigma^2}{\sqrt{2}\sigma}\right)\right) \qquad \mu > 0, \sigma > 0, \tau > 0 \tag{5.2}$$

This equation is also notated as $G_e(t;\mu,\sigma,\tau)$, which indicate that is the CDF of EMG as followers headway distribution.

The mean and variance of EMG distribution is:

$$E[X] = \mu + \frac{1}{\tau} \tag{5.3}$$

$$\mathrm{Var}(X) = \mu^2 + \frac{1}{\tau^2} \tag{5.4}$$

The Laplace transform of EMG distribution is:

$$g^*(s) = \frac{\exp\left(-\mu s + \dfrac{\sigma^2 s^2}{2}\right)}{1 + \dfrac{s}{\tau}} \tag{5.5}$$

which is also notated as $g_e^*(s)$.

Shifted-EMG distribution

By introducing the location parameter δ into Eq. 5.1 and 5.2, the PDF and CDF functions of shifted-EMG can be gained as follows:

PDF:

$$g(t;\mu,\sigma,\tau,\delta) = \frac{\tau}{2} e^{\tau\left(\mu - t + \delta + \frac{\tau\sigma^2}{2}\right)} \operatorname{erfc}\left(\frac{\mu - t + \delta + \tau\sigma^2}{\sqrt{2}\sigma}\right) \qquad \mu > 0, \sigma > 0, \tau > 0 \tag{5.6}$$

CDF:

$$G(t;\mu,\sigma,\tau,\delta)=\frac{1}{2}\left[\operatorname{erfc}\left(\frac{(\mu-t+\delta)}{\sqrt{2}\sigma}\right)-e^{\tau\left(\mu-t+\delta+\frac{\tau\sigma^2}{2}\right)}\operatorname{erfc}\left(\frac{\mu-t+\delta+\tau\sigma^2}{\sqrt{2}\sigma}\right)\right]$$
$$\mu>0,\sigma>0,\tau>0 \tag{5.7}$$

The shifted-EMG, in fact, has identical form of PDF and CDF functions as EMG distribution, since it is equivalent to replacing μ in EMG distribution with $\mu+\delta$. Because of this, the shifted-EMG should have the same performance as EMG distribution. However, this model is still tested as a separate model, which might show different results as both μ and δ will be treated separately using the numerical MLE method.

Inverse Gaussian(IVG) distribution

In the study, the IVG model is implemented using the built-in methods provided by Matlab. For the detailed mathematical properties can refer to book of Chhikara (1988) or Forbes et al. (2011). Heathcote (2004) gives details study on fitting IVG, shifted-IVG and ex-IVG distributions using response time data, and he provides mathematical details of these models in his paper.

PDF:

$$g(t;\mu,\sigma)=\sqrt{\frac{\sigma}{2\pi t^3}}\exp\left[-\frac{\sigma}{2\mu^2 t}(t-\mu)^2\right]$$
$$\mu>0,\sigma>0,t>0 \tag{5.8}$$

This equation is also notated as $g_i(t;\mu,\sigma)$, which indicate that is the PDF of IVG as followers headway distribution.

CDF:

$$\begin{aligned} G(t;\mu,\sigma) = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{\sigma}{2t}}\left(1-\frac{t}{\mu}\right)\right) \\ + \frac{1}{2}\exp\left(\frac{2\sigma}{\mu}\right)\operatorname{erfc}\left(\sqrt{\frac{\sigma}{2t}}\left(1+\frac{t}{\mu}\right)\right) \\ \mu > 0, \sigma > 0, t > 0 \end{aligned} \tag{5.9}$$

This equation function is also notated as $G_i(t;\mu,\sigma)$, which indicate that is the CDF of IVG as followers headway distribution.

The mean and variance of the IVG distribution is:

$$E[X] = \mu \tag{5.10}$$

$$\operatorname{Var}(X) = \frac{\mu^3}{\sigma} \tag{5.11}$$

The Laplace transform of the IVG distribution is (Lin, 1999):

$$g^*(s) = \exp\left\{\left(\frac{\sigma}{\mu}\right)\left(1-\sqrt{1+\frac{2\mu^2 s}{\sigma}}\right)\right\}, s > 0. \tag{5.12}$$

which is also notated as $g_i^*(s)$.

Shifted-IVG distribution

By introducing the location parameter δ into Eq. 5.8 and 5.9, the PDF and CDF functions of shifted-IVG can be gained as follows:

PDF:

$$g(t;\mu,\sigma,\delta) = \sqrt{\frac{\sigma}{2\pi(t-\delta)^3}}\exp\left[-\frac{\sigma}{2\mu^2(t-\delta)}(t-\delta-\mu)^2\right] \qquad \mu>0,\sigma>0,t>0 \tag{5.13}$$

CDF:

$$\begin{aligned} G(t;\mu,\sigma,\delta) &= \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{\sigma}{2(t-\delta)}}\left(1-\frac{t-\delta}{\mu}\right)\right) \\ &\quad + \frac{1}{2}\exp\left(\frac{2\sigma}{\mu}\right)\mathrm{erfc}\left(\sqrt{\frac{\sigma}{2(t-\delta)}}\left(1+\frac{t-\delta}{\mu}\right)\right) \\ &\quad \mu>0,\sigma>0,t>0 \end{aligned} \tag{5.14}$$

Note that the shifted-IVG distribution, also known as shifted-Wald, is itself a well-known and widely used RT model (Heathcote, 2004; Matzke and Wagenmakers, 2009).

5.1.3 Model Fitting to Headway Samples

The parameter estimation of EMG, IVG and shifted-IVG distributions are well studied for fitting RT models. Van Zandt (2000) examined varies estimation techniques on fitting RT models (including EMG and IVG models). He suggested that the MLE or least square estimators of CDF for fitting those models to the data.

Heathcote (2004) tested the MLE method on fitting ex-Wald (convolution of *exponential* and IVG distributions), EMG and shifted-IVG distributions using software package of S-PLUS. He suggested that MLE method gives consistent and small biased estimation of selected models when the test sample sizes are sufficient. He also reported equivalent fitting performance for both IVG and EMG models when they are tested in Monte Carlo simulation.

Silver et al. (2009) discussed estimation methods for fitting the EMG distribution, in the application of microarray background correction. He compared exact MLE, RMA, saddle-point estimates etc. He proposed an exact MLE method which will avoid the numerical sensitivity issues of the likelihood function (Bolstad, 2004; McGee and Chen, 2006). He found the exact MLE method return accurate estimates for EMG models comparing to the rest of methods.

An implementation level discussion of fitting EMG models using Matlab was provided by Lacouture and Cousineau (2008), where they used MLE to estimate the parameters of EMG models. He discussed the PDF of the EMG model and tested the MLE method using Monte Carlo study, and he found that MLE provides good parameter estimations for EMG distribution. The MLE method he used is the same as the numerical MLE method proposed in this thesis. In this study, Lacouture's method is used for the implementation of EMG distribution of headway analysis.

As one of the built-in models, the parameter estimation of IVG distribution uses MLE methods provided by MATLAB Statistical Toolbox.

Both shifted-EMG and shifted-IVG are estimated using the numerical MLE method similar to what has been discussed in chapter 4. Some tests have been done to compare shifted-IVG and IVG estimations, and the results show that when the location δ is set to zero, both methods will have identical results for parameter estimation. Similar tests are also performed to compare shifted-EMG and EMG estimation, and the results are again identical. These tests shows that numerical MLE is a valid method to fit these models to headway data.

All the 132 headway samples discussed in chapter 3 are fitted using the above methods. The results will be discussed in the following section.

5.1.4 Results and Discussion

For both EMG and IVG models, the estimation times are lower than 0.3s, while the fittings to IVG are slightly faster. The location parameter δ did not affect the fitting time much, they are very close to those fittings using the original models.

TABLE 5.1: Fitting time(s) on estimation of EMG and IVG distributions

	EMG		IVG	
	Mean	Std	Mean	Std
Shifted	0.23	0.05	0.16	0.01
Original	0.26	0.08	0.15	0.02

The fitting results are well accepted for samples of lane 1 and 2, but very poor in lane 3, as shown in Table 5.2, and Figures 5.1 and 5.2. There is no much difference with using location parameter δ in EMG models, as EMG and shifted-EMG fittings have same numbers of samples passed the GoF test. For IVG models, the location

Table 5.2: Summary of GoF result on EMG and IVG models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
EMG						
Shifted	46(6)	48(4)	43(4)	44(3)	0(33)	0(33)
Original	46(6)	48(4)	43(4)	44(3)	0(33)	0(33)
IVG						
Shifted	46(6)	48(4)	45(2)	47(0)	4(29)	6(27)
Original	34(18)	37(15)	24(23)	24(23)	0(33)	0(33)

parameter δ seems to have very positive effect, as a big improvement have been shown in the table, where the total number of samples passed GoF test increased from 58 to 91, in lane 1 and 2. The result of EMG and shifted-IVG have similar fitting performance compared to *shifted-lognormal* distribution tested in chapter 2. For the samples in lane 2, 2 samples have failed to gain parameters in fitting stages, while all the 47 samples have enabled parameters to be estimated with EMG and IVG models.

Figures 5.1 and 5.2 show more detailed goodness-of-fit results across traffic flows. The lane 3 shows very poor fitting results in all models, while lane 1 and 2 are fitted reasonable well. For the shifted-IVG and EMG (similar to shifted-EMG) models, more samples under relative higher traffic flows have mostly passed GoF test, while a few samples in lower traffic flow have failed.

For headway samples in lane 3, the fitting results are very poor, as non of EMG and IVG fittings passed the GoF test, and only a few passed GoF test (4 in KS test and 6 in AD test). These have been similar again to *shifted-lognormal* fittings. It seems that the single models have difficulty in describing headway data of lane

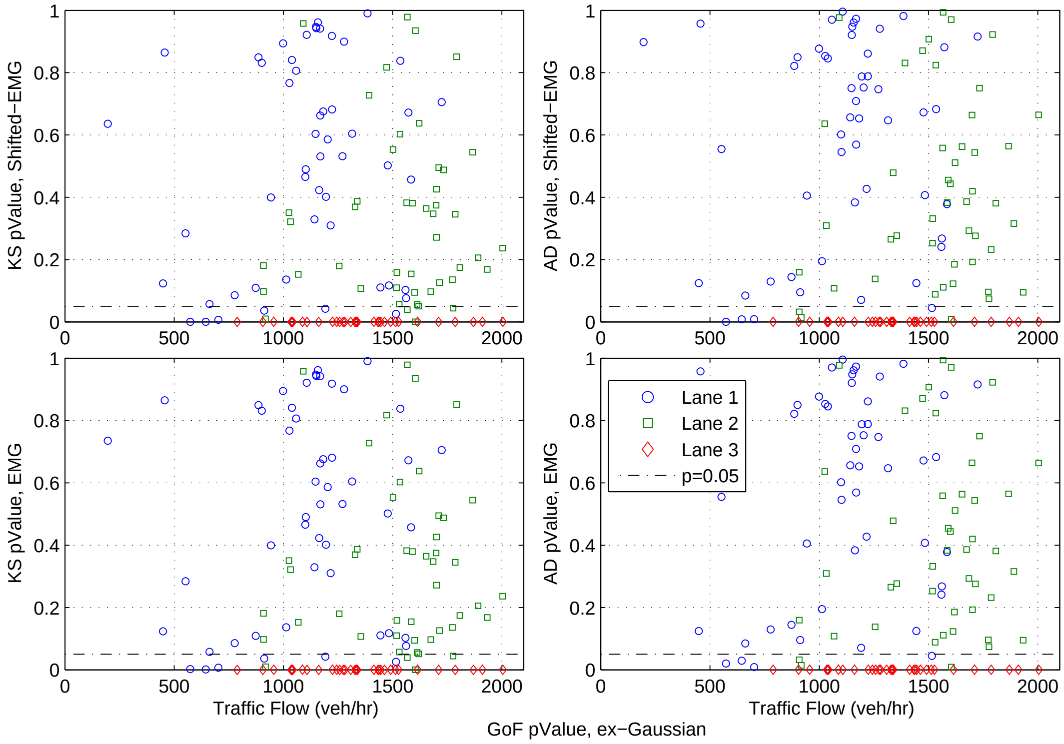

FIGURE 5.1: GoF results of EMG and shifted-EMG distributions

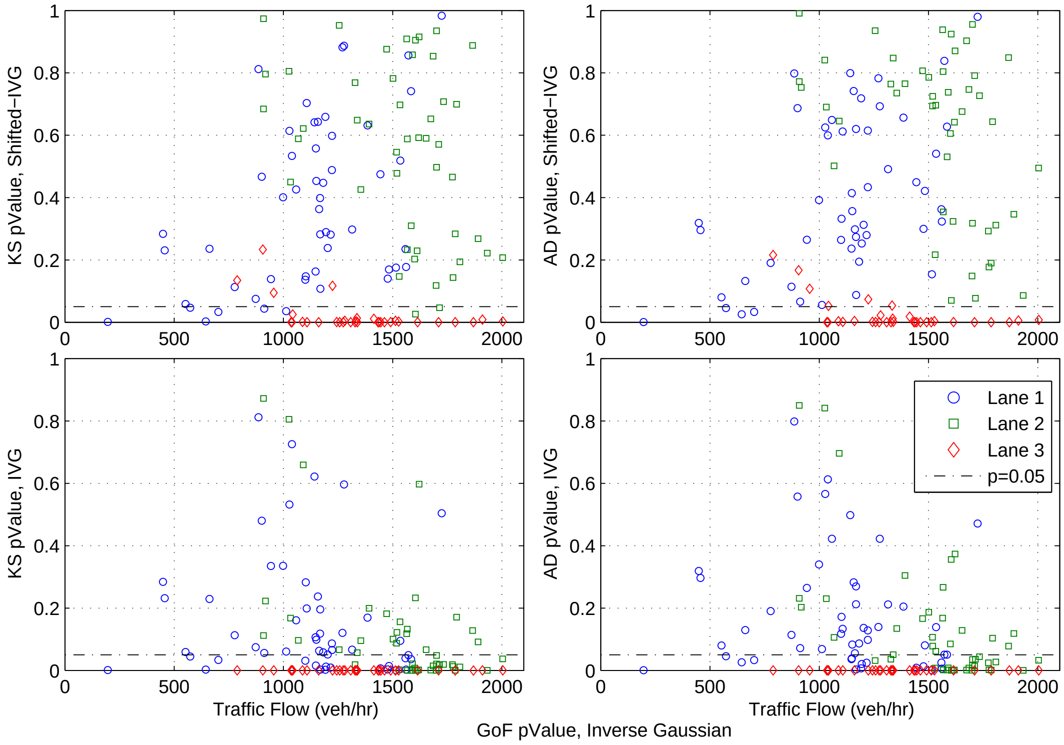

FIGURE 5.2: GoF results of IVG and shifted-IVG distributions

3.

Table 5.3 shows the estimated parameter δ, where all the δ values are very close to zero for EMG model. This indicate that the location parameter δ has not effect on the fitting performance of EMG model. As mentioned earlier, the EMG and shifted-EMG models have identical analytical forms. The very small δ can been seen as further evidence of consistency of parameter estimation with numerical MLE method. Note there is no data available for lane 3 in this table, as those values are calculated only using samples accepted by GoF test.

In comparison, the shifted-IVG has significant values of δ, which have improved the fitting result.

TABLE 5.3: Estimated δ in fitting with shifted-EMG and shifted-IVG models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
Shifted-EMG	0.001	0.003	0.002	0.001	n/a	n/a
Shifted-IVG	0.17	0.11	0.21	0.06	0.31	0.05

Figure 5.3 illustrates the estimated PDF curves in comparison with the histogram of headway data of 6 selected samples. In the figure, the EMG and shifted-EMG models look almost identical, and they provide a good fit to samples lane 1 and lane 2 (the top four charts). The PDF curves of IVG model are apparently different from shifted-IVG model. The shifted-IVG shows good fitting visually in the figures, as well do the EMG and shifted-EMG models. For the bottom two charts of samples in lane 3, all the PDF curves do not seem to fit the histogram well.

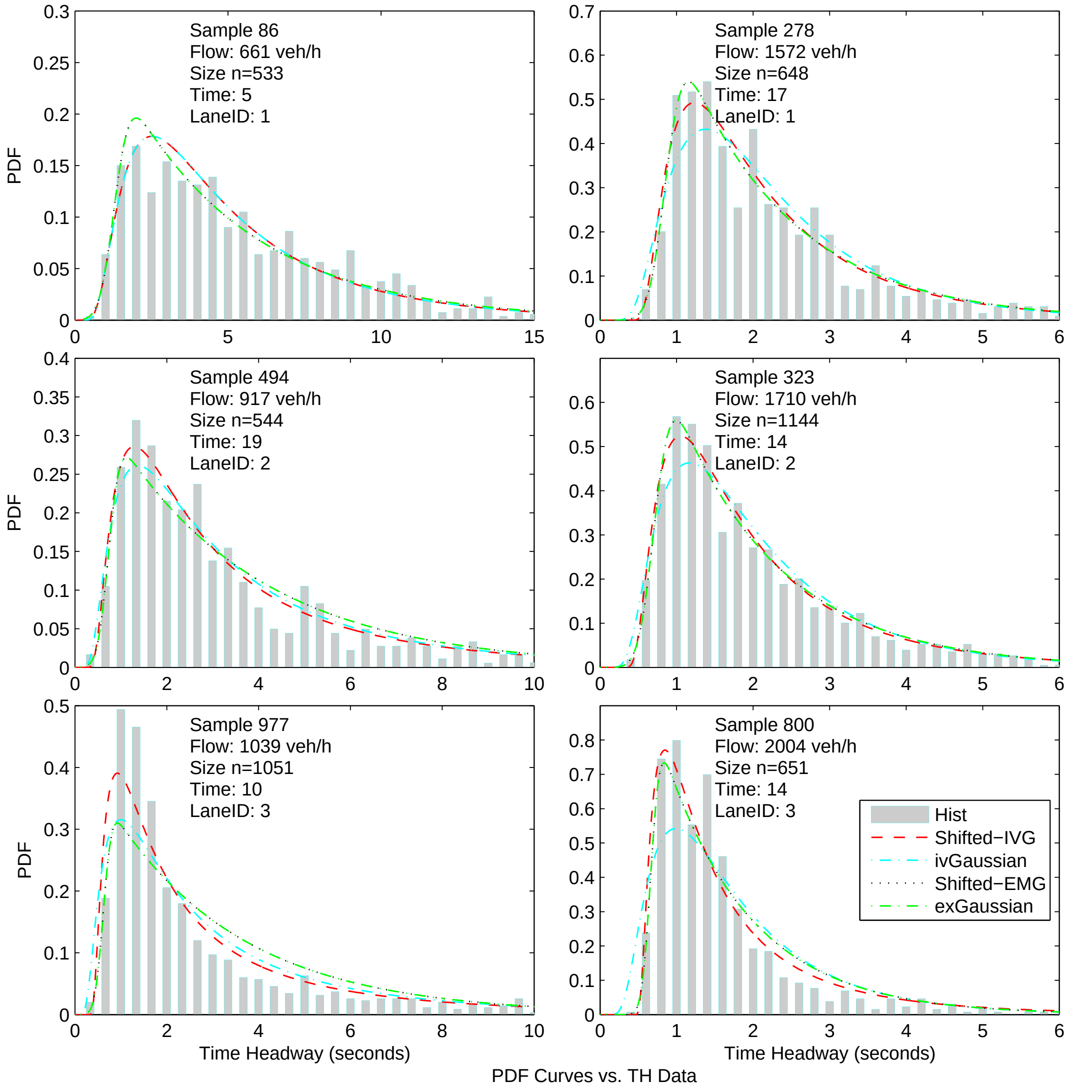

FIGURE 5.3: PDF fitting results of EMG, shifted-EMG, IVG and shifted-IVG on selected data samples

Overall, as single models, both shifted-IVG and EMG have good fitting performance under some of traffic conditions, and their fitting results are comparable with the *shifted-lognormal* distribution. These results are very promising, as the two models are chosen from widely used RT models, which are, for the first time, tested with headway data. As a positive example, the above test results certainly indicate some potential for using RT models in headway analysis. Just as with the other single models, even though the test results are very positive for lane 1 and

2, these models are also have difficulties for describing the headway under more complicated traffic situation, in lane 3.

To extend this, both EMG and IVG models are further attempted as followers headway distributions, and the next section investigates how these two RT models can behave when they are used in SPM and GQM mixed models.

5.2 RT Models as Followers Headway Distribution

The fitting results show reasonable fitting performance to headway data for both IVG and EMG models in the above section. These results provide more confidence of using such RT models in headway analysis. As discussed in chapter 4, using mixed distribution models can greatly improve fitting results when separating headway data into non-followers and followers group. This might be explained by that the selected single distributions (*gamma* and *lognormal*) are more capable of describing headway data in car following conditions, i.e. when drivers have interactions with their preceding vehicles. In other words, these followers are more prepared for a change compared with the non-followers. This sounds analogous to some of those tasks that measure RT times, while subjects are prepared for a change of stimuli before a decision is made for their next action (e.g. waiting to hear a sound before braking (Johansson and Rumar, 1971)). This gives an intuition that using RT models to describe followers headway in SPM and GQM mixed models may also fit headway data well.

This section introduces four mixed models using the selected RT models (IVG and EMG) as followers headway distributions, i.e. EMG-SPM, EMG-GQM, IVG-SPM and IVG-GQM models. These models are implemented within the proposed framework of the present study, with discussion of their fitting performance. As in chapter 4, some parameter simplification methods are also discussed for these four mixed models.

5.2.1 EMG and IVG Distribution in Mixed Models

For SPM mixed models, the general form PDF and CDF are shown in Eq. 4.6 and 4.7; for GQM mixed models, please refer to Eq. 4.10 and 4.12. Substituting $g(t)$ and $G(t)$ with equations of EMG and IVG distributions, the function form of EMG-mixed and IVG-mixed models can be gained.

Since the numerical MLE method estimates parameters directly using the calculation of PDF in likelihood function of each model, the equations of PDF and CDF are not required for further expansion hence simplification, and they will remain in the form similar to the general equations of SPM and GQM models. This can simplify the implementation of each model, as it can maximize the reuse of the Matlab codes of PDF and CDF of EMG and GQM that are already developed in the earlier part of this chapter. As described in chapter 2, this is one of the advantages of using the proposed framework of the present study. The following few sections describe the PDF and CDF for each proposed mixed model, and for the GQM models, some simplification of equations are introduced while the functions still remain in the same form.

EMG-SPM

The PDF and CDF of EMG-SPM model can be achieved by substituting Eq. 5.1, 5.2 and 5.5 into Eq. 4.6 and 4.7.

PDF:

$$f(t;\phi,\lambda,\mu,\sigma,\tau)=\phi g_e(t;\mu,\sigma,\tau)+(1-\phi)\frac{G_e(t;\mu,\sigma,\tau)}{g_e^*(\lambda)}\lambda e^{-\lambda t}, \quad t\geq 0;\lambda>0;0\leq\phi\leq 1 \tag{5.15}$$

CDF:

$$F(t;\phi,\lambda,\mu,\sigma,\tau)=\phi G_e(t;\mu,\sigma,\tau)+\frac{1-\phi}{g_e^*(\lambda)}\int_0^t G_e(u;\mu,\sigma,\tau)\lambda e^{-\lambda u}\mathrm{d}u, \quad t\geq 0;\lambda>0;0\leq\phi\leq 1 \tag{5.16}$$

EMG-GQM

The PDF and CDF of EMG-GQM model can be achieved by substituting Eq. 5.1, and 5.2 into Eq. 4.10 and 4.12.

PDF:

$$f(t;\phi,\lambda,\mu,\sigma,\tau)=\phi g_e(t;\mu,\sigma,\tau)+(1-\phi)\lambda e^{-\lambda t}\int_0^t g_e(u;\mu,\sigma,\tau)e^{\lambda u}\mathrm{d}u \tag{5.17}$$

CDF:

$$F(t;\phi,\lambda,\mu,\sigma,\tau)=G_e(t;\mu,\sigma,\tau)-(1-\phi)e^{-\lambda t}\int_0^t g_e(u;\mu,\sigma,\tau)e^{\lambda u}\mathrm{d}u \tag{5.18}$$

The above two equations can be further derived, by simplifying the integration part of those equations.

$$\int_0^t g_e(u;\mu,\sigma,\tau)e^{\lambda u}\mathrm{d}u = g_e^*(-\lambda)G_e(t;\mu+\lambda\sigma^2,\sigma,\frac{1}{\tau-\lambda})$$

Hence the simplified PDF and CDF are:

PDF:

$$f(t;\phi,\lambda,\mu,\sigma,\tau) = \phi g_e(t;\mu,\sigma,\tau) + (1-\phi)\lambda g_e^*(-\lambda)G_e(t;\mu+\lambda\sigma^2,\sigma,\frac{1}{\tau-\lambda})e^{-\lambda t} \tag{5.19}$$

CDF:

$$F(t;\phi,\lambda,\mu,\sigma,\tau) = G_e(t;\mu,\sigma,\tau) - (1-\phi)g_e^*(-\lambda)G_e(t;\mu+\lambda\sigma^2,\sigma,\frac{1}{\tau-\lambda})e^{-\lambda t} \tag{5.20}$$

where $g_i^*(-\lambda)$ is the Laplace transform of PDF of EMG, with value $-\lambda$, which is:

$$f^*(-\lambda) = \frac{\exp\left(\mu\lambda + \dfrac{\sigma^2\lambda^2}{2}\right)}{1-\dfrac{\lambda}{\tau}} \tag{5.21}$$

For the above simplification to be valid, the parameters has to satisfy the condition that $\tau \neq \lambda$.

IVG-SPM

The PDF and CDF of IVG-SPM model can be achieved by substituting Eq. 5.8, 5.9 and 5.12 into Eq. 4.6 and 4.7.

PDF:

$$f(t;\phi,\lambda,\mu,\sigma) = \phi g_i(t;\mu,\sigma) + (1-\phi)\frac{G_i(t;\mu,\sigma)}{g_i^*(\lambda)}\lambda e^{-\lambda t}, \qquad t \geq 0; \lambda > 0; 0 \leq \phi \leq 1 \tag{5.22}$$

CDF:

$$F(t;\phi,\lambda,\mu,\sigma) = \phi G_i(t;\mu,\sigma) + \frac{1-\phi}{g_i^*(\lambda)}\int_0^t G_e(u;\mu,\sigma)\lambda e^{-\lambda u}\mathrm{d}u, \qquad t \geq 0; \lambda > 0; 0 \leq \phi \leq 1 \tag{5.23}$$

IVG-GQM

The PDF and CDF of IVG-GQM model can be achieved by substituting Eq. 5.8, and 5.9 into Eq. 4.10 and 4.12.

PDF:

$$f(t;\phi,\lambda,\mu,\sigma) = \phi g_i(t;\mu,\sigma) + (1-\phi)\lambda e^{-\lambda t}\int_0^t g_i(u;\mu,\sigma)e^{\lambda u}\mathrm{d}u \tag{5.24}$$

CDF:

$$F(t;\phi,\lambda,\mu,\sigma) = G_i(t;\mu,\sigma) - (1-\phi)e^{-\lambda t}\int_0^t g_i(u;\mu,\sigma)e^{\lambda u}\mathrm{d}u \tag{5.25}$$

To avoid the integration of $g_i(t)$ for above two equations, the integration part can be further derived and hence simplified using:

$$\int_0^t g_i(u;\mu,\sigma)e^{\lambda u}\mathrm{d}u = g_i^*(-\lambda)G_i(t;\frac{\mu}{\sqrt{1-\frac{2\lambda\mu^2}{\sigma}}},\sigma)$$

Hence the simplified PDF and CDF are:

PDF:

$$f(t;\phi,\lambda,\mu,\sigma) = \phi g_i(t;\mu,\sigma) + (1-\phi)\lambda g_i^*(-\lambda)G_i(t;\frac{\mu}{\sqrt{1-\frac{2\lambda\mu^2}{\sigma}}},\sigma)e^{-\lambda t} \tag{5.26}$$

CDF:

$$F(t;\phi,\lambda,\mu,\sigma) = G_i(t;\mu,\sigma) - (1-\phi)g_i^*(-\lambda)G_i(t;\frac{\mu}{\sqrt{1-\frac{2\lambda\mu^2}{\sigma}}},\sigma)e^{-\lambda t} \tag{5.27}$$

where $g_i^*(-\lambda)$ is the Laplace transform of PDF of IVG, with value $-\lambda$, which is:

$$g_i^*(-\lambda) = \exp\left\{\left(\frac{\sigma}{\mu}\right)\left(1-\sqrt{1-\frac{2\lambda\mu^2}{\sigma}}\right)\right\} \tag{5.28}$$

This simplification can greatly reduce the computation when PDF and CDF of IVG are calculated. However, it is important to note, this can only work when $\sqrt{1-\frac{2\lambda\mu^2}{\sigma}}$ and $\frac{\mu}{\sqrt{1-\frac{2\lambda\mu^2}{\sigma}}}$ is valid, i.e. the parameters must always satisfy $\sigma > 2\lambda\mu^2$. In reality, this would not always happen when fitting with headway samples. As a result, this simplification has very limited usage. In the present study, the simplification is only used when the conditions are met. For the remaining estimations, numerical integration has to be used for computation.

5.2.2 Implementation of Mixed Models in MATLAB

All the above 4 mixed models are implemented using the MATLAB Statistical Toolbox. Since it is difficult or impossible to derive explicit forms for these distribution functions, these models can only be implemented numerically.

These models are again implemented using the proposed framework of this study. With the existence of the IVG and EMG distribution classes, the development of these four proposed models become very simple and straightforward. Similar procedures can be found in chapter 4.

All parameters are estimated using the numerical MLE method. This estimation method may not be optimum in terms of computation performance, however it is adequate for testing the fitting performance of the above proposed models. When testing the fitting results, both KS and AD GoF tests are used similar to the tests of the previous models.

A few aspects of the numerical MLE procedure for these models need special attentions.

In the models involve EMG distribution, there is numerical sensitivity issues using Eq. 5.1 as $\mathrm{erfc}(x)$ is subject to subtractive cancellation when $\dfrac{\mu - t + \tau\sigma^2}{\sqrt{2}\sigma}$ is too large, which in result will result $\mathrm{erfc}\left(\dfrac{\mu - t + \tau\sigma^2}{\sqrt{2}\sigma}\right)$ very small, e.g. $\mathrm{erfc}(35) <$

10^{-530} which would underflow at many machine precisions (Haney, 2011). This issue is considered by skipping those parameters that will lead to such problem.

When dealing with EMG-GQM model, Eq. 5.19 is used in estimation stage, as it can avoid the numerical integration which will be computationally costly. For the same reason, Eq. 5.26 is used for fitting IVG-GQM model. However, this equation can only be applied under limited conditions as discussed previously, hence it is partially used only when the condition is met. For the rest of situations, Eq. 5.24 is still in use.

Since the CDF of both EMG-SPM and IVG-SPM are very difficult, if not impossible, to be simplified into explicit function forms to avoid the numerical integrations, there will be higher computational cost in GoF tests comparing to GQM models..

In terms of setting initial values, similar methods are used that have been discussed in chapter 4.

All the 132 headway samples described in chapter 3 were used to fit the 4 mixed models.

5.2.3 Results and Analysis

Table 5.4 shows the average fitting time of the four proposed models. For the EMG-SPM, EMG-GQM and IVG-SPM models, all the fitting times are well below 1 second, which are lower than the fitting time of gamma-SPM and gamma-GQM models (see Table 4.1). As expected, the fitting time of IVG-GQM is higher with larger variance. This is due to part of estimations have to use Eq. 5.24 which involves numerical integration.

Table 5.4: Comparison of estimation and GoF time on EMG-SPM, EMG-GQM, IVG-SPM and IVG-GQM models

	Fitting Time(s)		KS test time(ms)		AD test time(ms)	
	Mean	Std	Mean	Std	Mean	Std
IVG-						
SPM	0.37	0.15	357	117	376	156
GQM	3.94	12.47	27	83	25	79
EMG-						
SPM	0.67	0.22	361	119	371	159
GQM	0.66	0.25	4	4	3	1

For the GoF test, both EMG-GQM and IVG-GQM show very good GoF test speeds, which are similar comparing with gamma-SPM and gamma-GQM models. In between, the EMG-GQM have showed the fastest and most stable speed of only a few milliseconds in GoF test. However, the speeds of GoF test with both EMG-SPM and IVG-SPM are much slower with greater variations, which are caused by the inevitable numerical integrations involved in computation of their CDFs.

Table 5.5: Summary of goodness-of-fit results on EMG-SPM, EMG-GQM, IVG-SPM and IVG-GQM models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
IVG-						
SPM	52(0)	52(0)	47(0)	47(0)	33(0)	33(0)
GQM	52(0)	52(0)	47(0)	47(0)	33(0)	33(0)
EMG-						
SPM	52(0)	52(0)	47(0)	47(0)	33(0)	33(0)
GQM	52(0)	52(0)	47(0)	47(0)	33(0)	33(0)

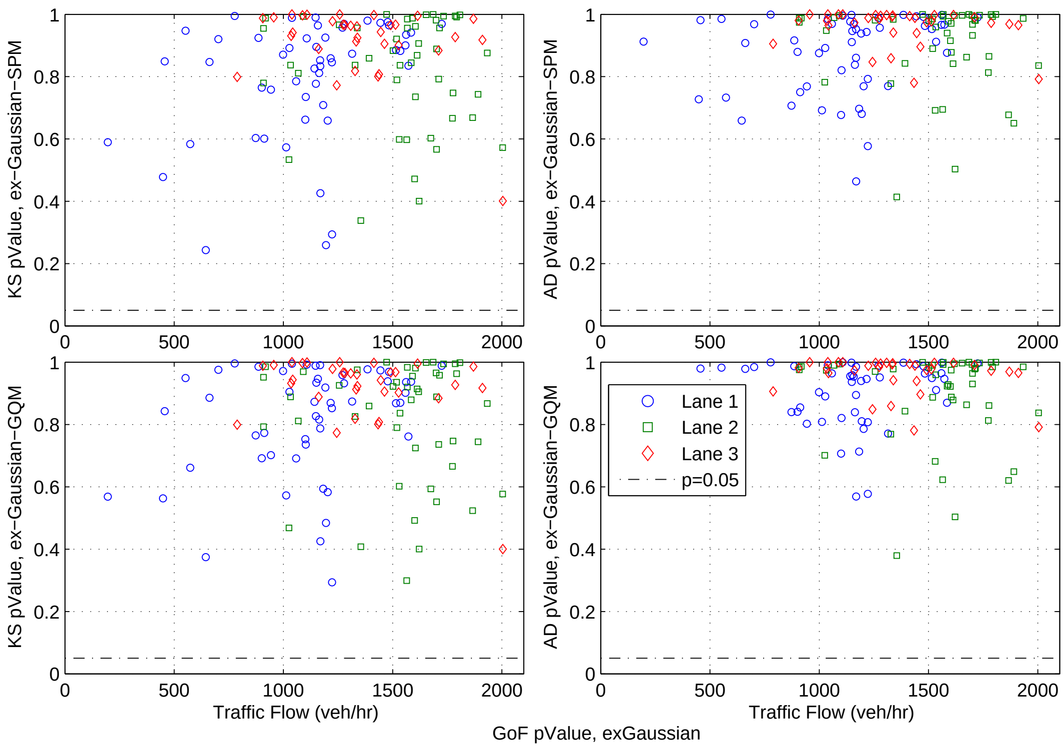

FIGURE 5.4: GoF result (p value) for EMG-SPM and EMG-GQM models

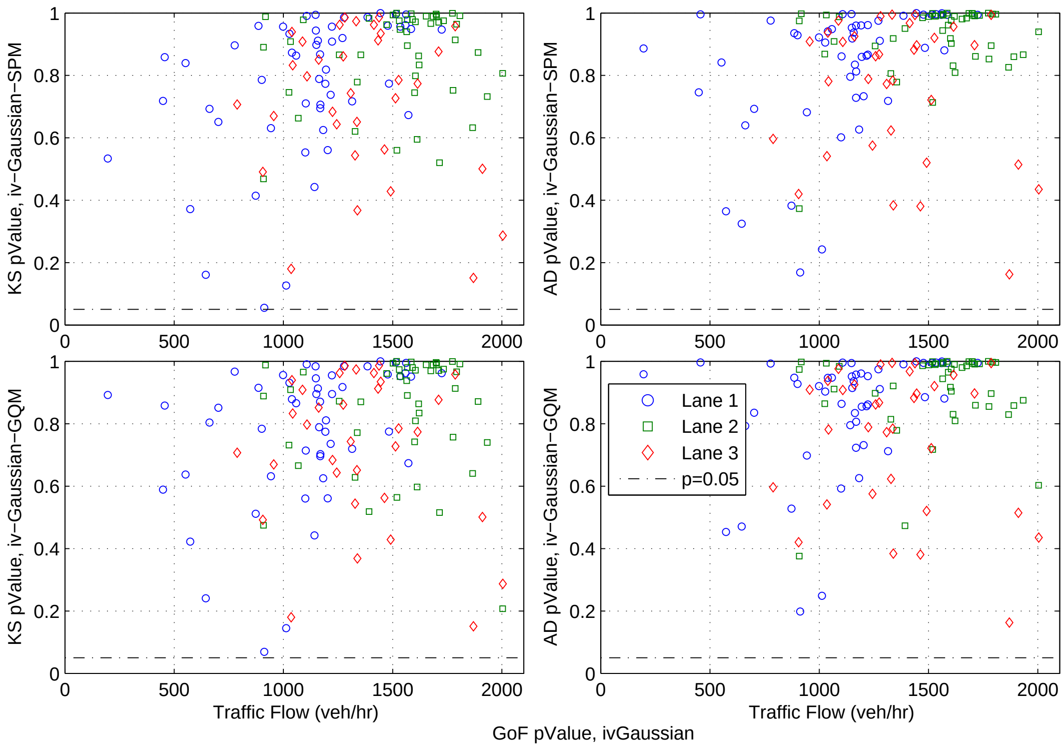

FIGURE 5.5: GoF result (p value) for IVG-SPM and IVG-GQM

The fitting performance is tested by using the GoF tests, with same methods as in Chapter 4. Table 5.5 presents the result of the GoF tests. These summarized results show that both EMG-mixed and IVG-mixed models are excellently accepted, as all headway samples have passed the GoF tests. These results are slightly better than the well fitted results for the gamma-mixed and lognormal-mixed models studied in Chapter 4 (see Table 4.2). Similar findings are also shown in Figures 5.4 and 5.5, in which most samples have higher p-values compared with the results in Figure 4.2 and 4.3. That is, not only are more samples accepted by using the proposed RT models as followers headway distribution, but also the samples are better fitted with these models. Higher p-values can be observed in the test results of the EMG-mixed models compared with the IVG-mixed models and this may be caused by greater flexibility in parameter estimations of 5-parameter EMG-mixed models. There is no clear difference observed between the results of the SPM and GQM models.

Figure 5.6 shows the visual fitting results for six selected samples, using the PDF of each model to compare with the histograms of the sampled headway data. The three rows in the figure present samples from lane 1-3 of the tested road, from top to bottom respectively. The left column shows samples with relatively lower traffic flow rates, while the right column shows samples with higher flow rates.

All the figures present the PDF curves of all four mixed models. Between the SPM and GQM models of either EMG or IVG, it is hardly to see any difference as the corresponding PDF curves are almost overlapped to each other. All the curves

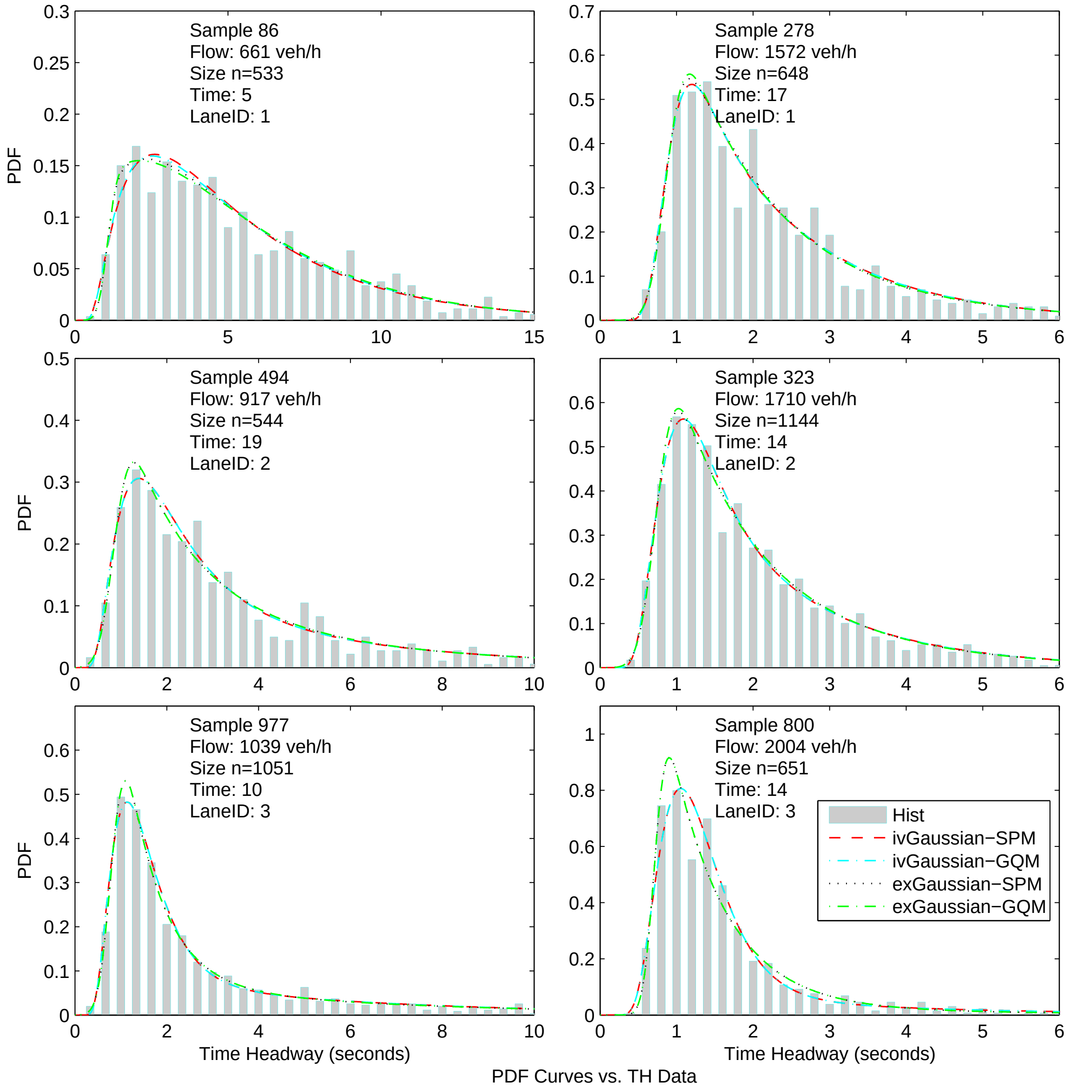

FIGURE 5.6: PDF of selected headway samples

seem to have good fittings to the histograms of the empirical headway data. These again show the similarities in fitting results between SPM and GQM models.

Between the EMG-mixed and IVG-mixed models, it is hardly to compare which has better fitting to the data as their curves appear similar.

The estimated parameters of all the four models are summarized in Table 5.6, and more details can be viewed visually in Figures 5.7-5.10. Comparing with the results of gamma-mixed and lognormal-mixed models (see Figures 4.5-4.8), the estimated values of the arrival rates parameter λ have very similar increasing trends with the increased level of traffic flows.

Table 5.6: Parameters variation on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
IVG-SPM						
ϕ	0.26	0.19	0.45	0.18	0.62	0.09
λ	0.43	0.13	0.53	0.15	0.27	0.06
μ	1.84	1.11	1.43	0.36	1.26	0.15
σ	10.20	2.69	8.25	1.32	8.45	1.75
IVG-GQM						
ϕ	0.31	0.18	0.52	0.20	0.62	0.09
λ	0.43	0.13	0.53	0.16	0.27	0.06
μ	1.88	1.22	1.49	0.41	1.26	0.15
σ	10.60	2.72	8.31	1.46	8.45	1.75
EMG-SPM						
ϕ	0.33	0.27	0.66	0.22	0.67	0.09
λ	0.43	0.13	0.46	0.16	0.24	0.05
μ	0.89	0.13	0.78	0.07	0.72	0.06
σ	0.20	0.08	0.19	0.04	0.16	0.02
τ	1.23	1.18	1.03	0.47	0.65	0.17
EMG-GQM						
ϕ	0.45	0.26	0.78	0.18	0.68	0.09
λ	0.46	0.13	0.44	0.15	0.24	0.05
μ	0.93	0.12	0.79	0.07	0.72	0.06
σ	0.22	0.08	0.20	0.04	0.16	0.02
τ	1.27	0.86	1.15	0.45	0.66	0.17

With the IVG-mixed models, μ and σ are the two parameters of the followers headway distribution. The values of parameter σ spread between 5 and 20, with most values between 7 and 15. These values of σ do not seem to follow any trend of changes in the traffic flows, and the values in lane 1 seem slightly higher than lane 2 and 3. Most of values of μ in IVG-mixed models are between 1 and 2, and there are no clear distinctions between all three lanes. In lane 1, values of μ are more scattered at lower traffic flow levels, and they show a decreasing trend with

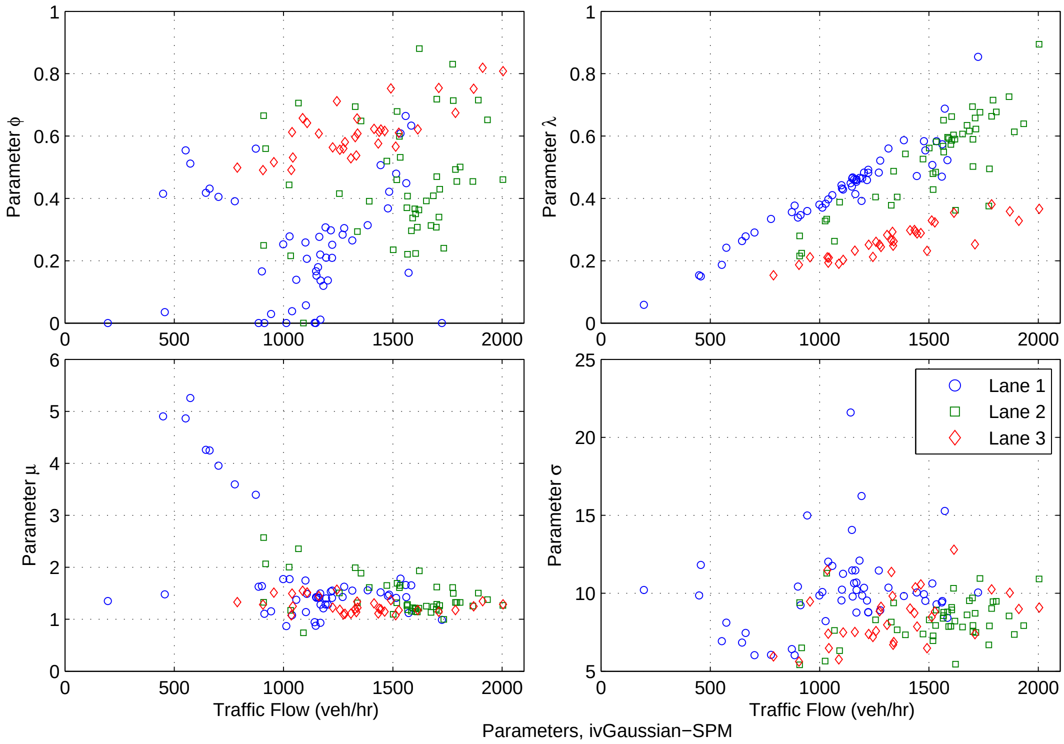

FIGURE 5.7: Estimated parameters for IVG-SPM

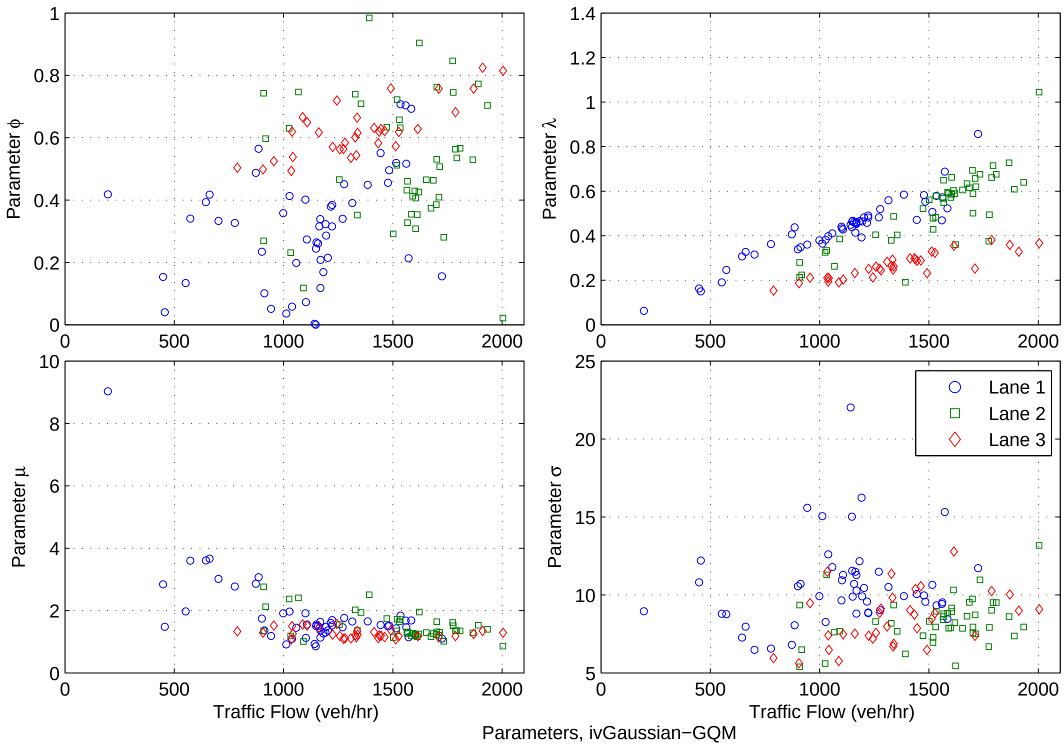

FIGURE 5.8: Estimated parameters for IVG-GQM

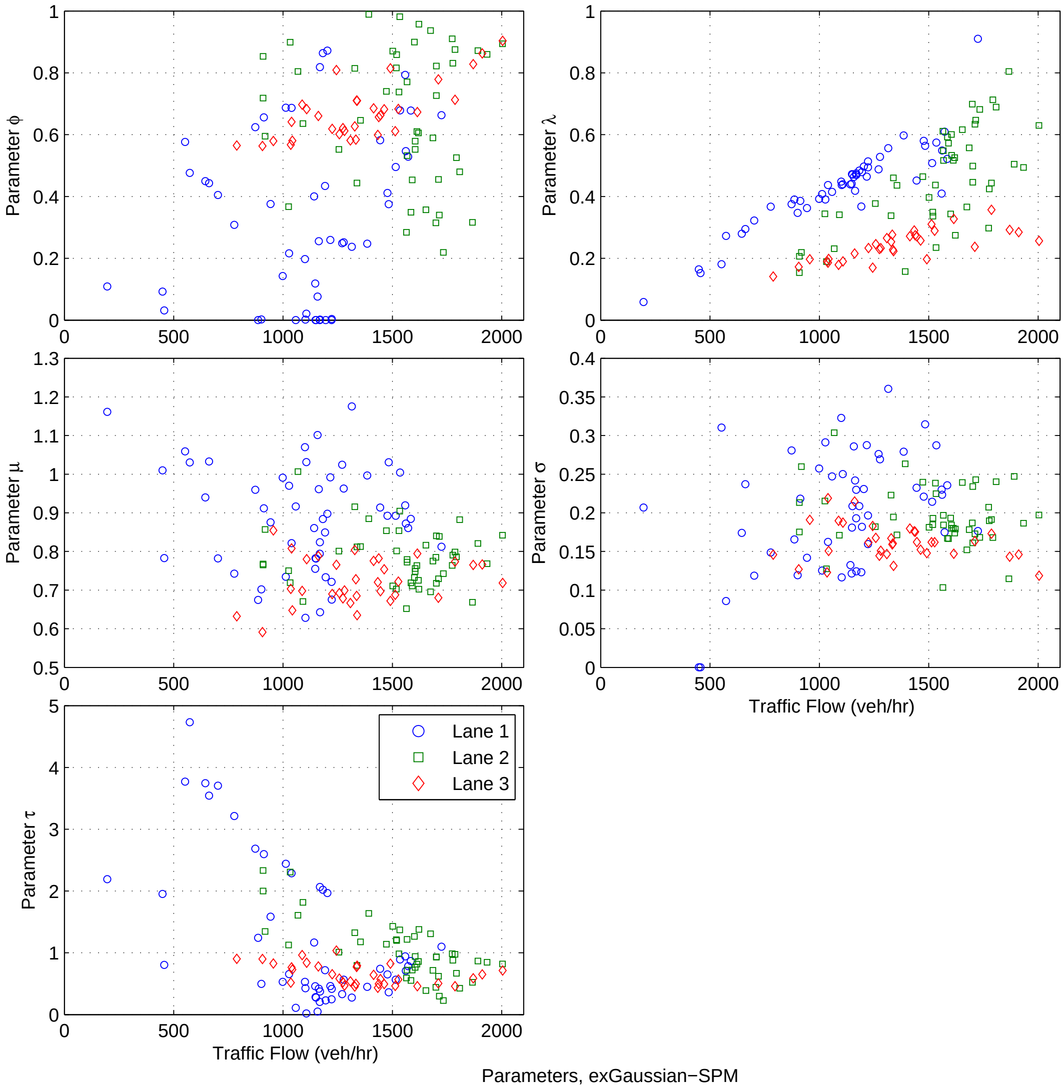

FIGURE 5.9: Estimated parameters for EMG-SPM

the increase of traffic flows. In general, values of both μ and σ have less variation when the traffic becomes heavier.

With the EMG-mixed models, the three parameters, μ, σ and τ seem to converge with the increased level of flows. The values of μ are gathered between 0.6 and 1.2, while values of σ are gathered between 0.1 and 0.4. The values of τ in lane 1 and 2 are more scattered in the lower traffic flow, while values in lane 3 seem more consistent.

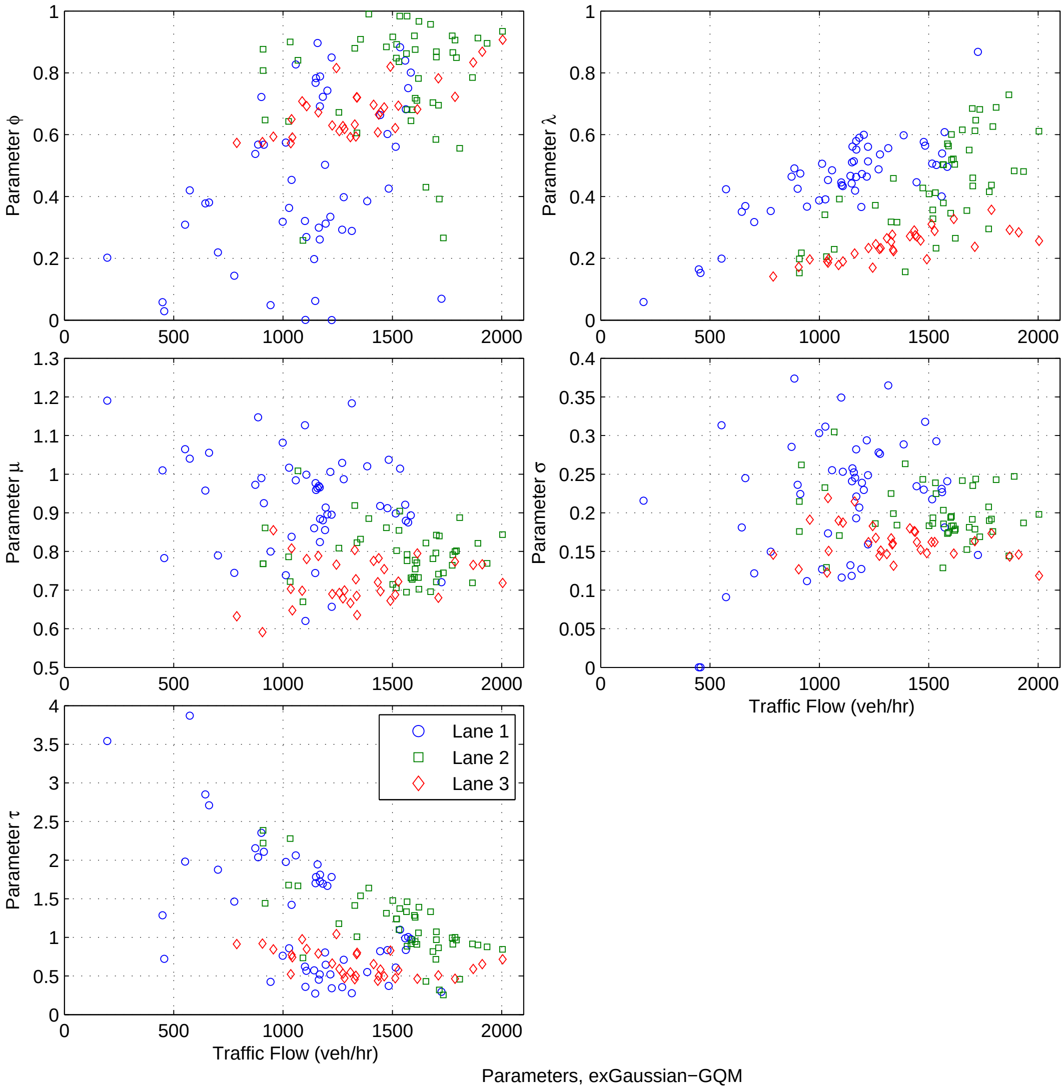

FIGURE 5.10: Estimated parameters for EMG-GQM

The general impressions for values of proportion parameter (ϕ) are very similar to what occurred in the gamma-mixed and lognormal-mixed models, especially for headways in lane 2 and 3. Some of those problematic estimations of parameter ϕ, as mentioned in chapter 4, seem to be improved, as they depart from zero. However, these values of ϕ seem higher than those ϕ values with higher levels of flows in the same lane, which is still unreasonable considering that there should be a smaller portion of followers in queues with lower traffic flows. The estimated values of ϕ with the two GQM models seem less problematic compared with those

ϕ values of SPM models. These again may be caused by the problem of over flexibility with 4- or 5-parameter models as discussed in chapter 4.

TABLE 5.7: Mean followers headway (s)

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
IVG-						
SPM	1.84	1.11	1.43	0.36	1.26	0.15
GQM	1.88	1.22	1.49	0.41	1.26	0.15
EMG-						
SPM	2.12	1.20	1.81	0.48	1.37	0.17
GQM	2.20	0.90	1.94	0.46	1.38	0.17

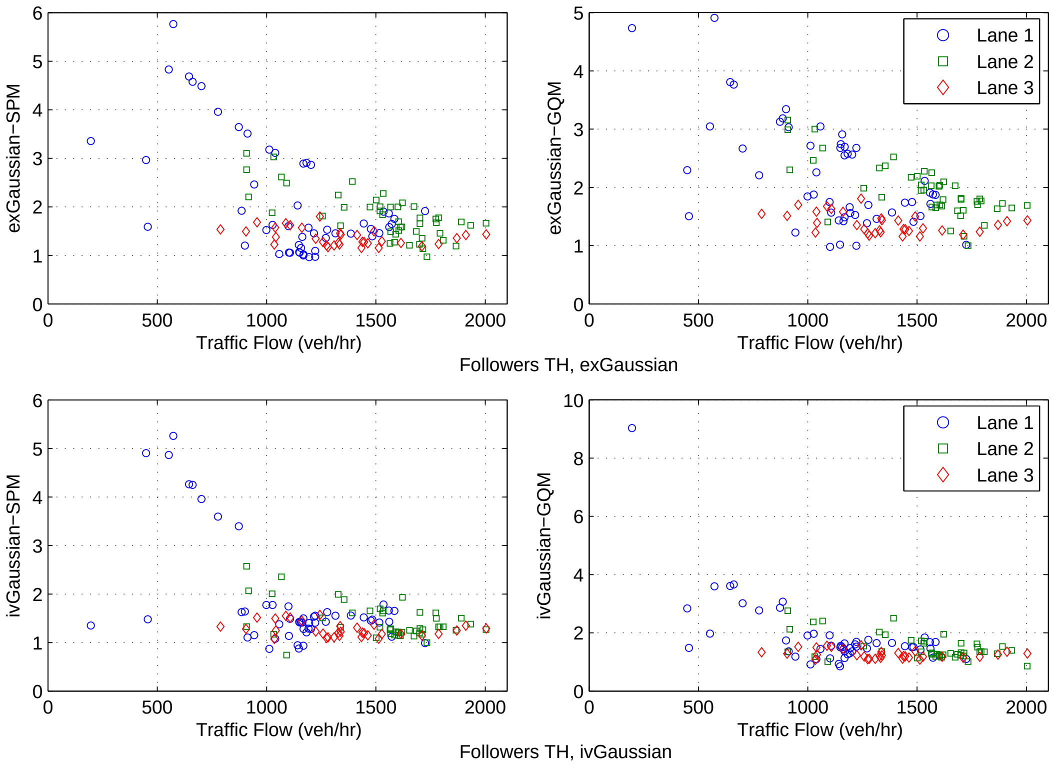

FIGURE 5.11: Mean followers headway (s) against traffic flows. The y-axis show the mean followers headway of each specified model.

The results for the mean followers headway are relatively consistent (shown in Table 5.7), and they are more aggregated when the traffic flow gets higher (Figure 5.11), within the scope between 0.8 and 3 seconds. Lane 1 has the largest and more scattered mean headway values, while lanes 2 and 3 have smaller variations with slightly higher mean headway values in lane 2. Similar trends are found and discussed in chapter 4 for the gamma-mixed and lognormal-mixed models (see Figure 4.9). The large variations in lane 1 may be due to those problematic estimations mentioned previously in this section, and these may potentially be improved with more constraints applied to the models for simplification.

The estimated mean followers headways of EMG-mixed models are higher than those of IVG-mixed models, across all the three lanes, while both EMG-mixed and IVG-mixed models have slightly higher mean headway values compared with gamma-mixed and lognormal-mixed models. There are strong similarities between SPM and GQM models, for all the four models.

Overall, all of the proposed EMG-mixed and IVG-mixed models have very good fitting performance and reasonable computation cost on parameter estimation. The tests shows generally better fitting performances compared with the gamma-mixed or lognormal-mixed models.

As an attempt at using selected RT models to describe followers headway distributions, the above implementations show very promising results which strongly support the assumptions of this chapter, i.e. both EMG and IVG distributions can be good alternative options in headway analysis.

5.2.4 Methods of Simplification on Estimation of Models

In the next few sections, some steps, similar to chapter 4, are used as attempts of simplifying the estimation process of all the four mixed models studied in this chapter. Firstly, attention is paid to the pre-determination of parameter λ. Following with some tests using combination of parameter λ and selected parameters of the followers' distribution. To save some space, some results are only shown in the tables in the main chapter, and related figures can be found in Appendix C.

5.2.5 Methods of Pre-determining Parameter λ

The simplifications for EMG-mixed and IVG-mixed models are very similar to those discussed in Chapter 4. The same equations (in Table 4.5) are used here to pre-determine λ, and similar estimation methods are applied.

TABLE 5.8: Comparison of estimation and GoF time on EMG-SPM, EMG-GQM, IVG-SPM and IVG-GQM models

	Fitting Time(s)		KS test time(ms)		AD test time(ms)	
	Mean	Std	Mean	Std	Mean	Std
IVG-						
SPM	0.26	0.05	346	112	362	158
GQM	0.79	3.71	9	38	8	39
EMG-						
SPM	0.42	0.12	362	120	378	165
GQM	0.44	0.15	4	3	3	1

The fitting speeds are, as expected, improved since one parameter reduced from the estimations, and the fitting time can be found in Table 5.8.

Table 5.9 provides with summarized fitting results, which show a well acceptance by GoF tests, especially for the data of lanes 2 and 3. The results in lane 1 are slightly degraded compared to those in table 5.5, with 2-4 samples failed during the tests. The overall results are very similar compared to the gamma-mixed models

TABLE 5.9: GoF results with pre-determined λ on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
IVG-						
SPM	48(4)	52(0)	47(0)	47(0)	33(0)	33(0)
GQM	49(3)	50(2)	47(0)	47(0)	33(0)	33(0)
EMG-						
SPM	50(2)	51(1)	47(0)	47(0)	33(0)	33(0)
GQM	50(2)	51(1)	47(0)	47(0)	33(0)	33(0)

The numbers with or without brackets show the numbers of successes or failures of GoF test.

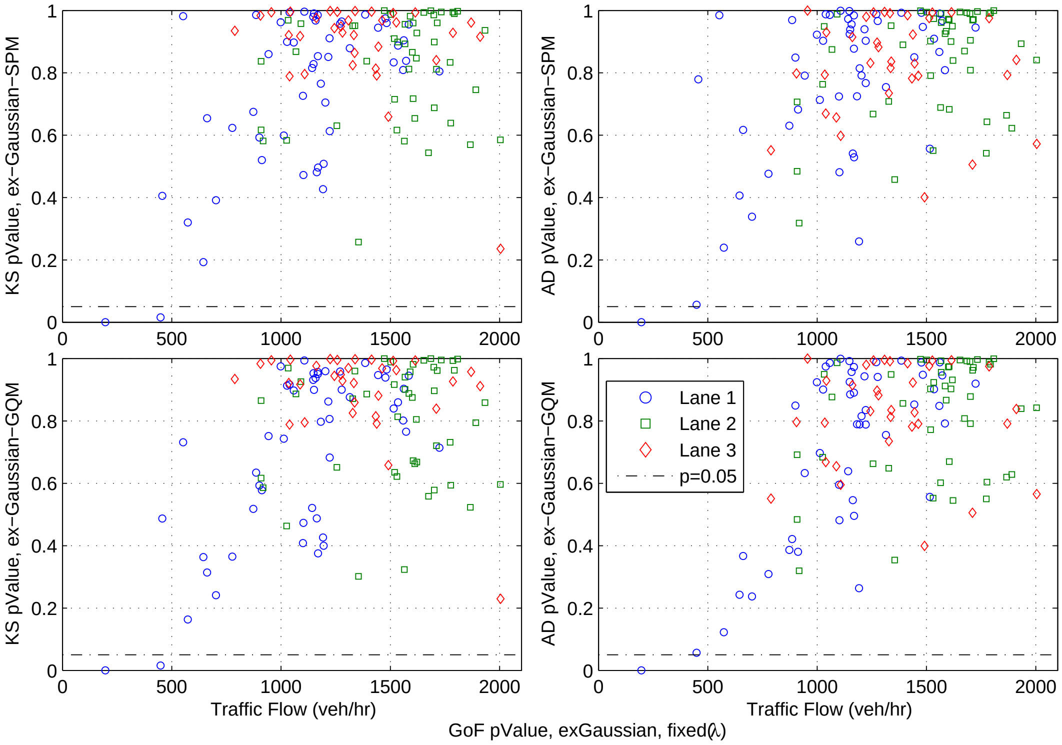

FIGURE 5.12: GoF result (p value) for EMG-SPM and EMG-GQM models with pre-determined λ

(see Table 4.6). The Figures 5.12 and 5.13 show averagely higher p-values for both EMG-mixed and IVG-mixed models, which indicate generally better fitting performance, compared to results of gamma-mixed models.

With the pre-determined λ, the estimated parameters have not been greatly changed,

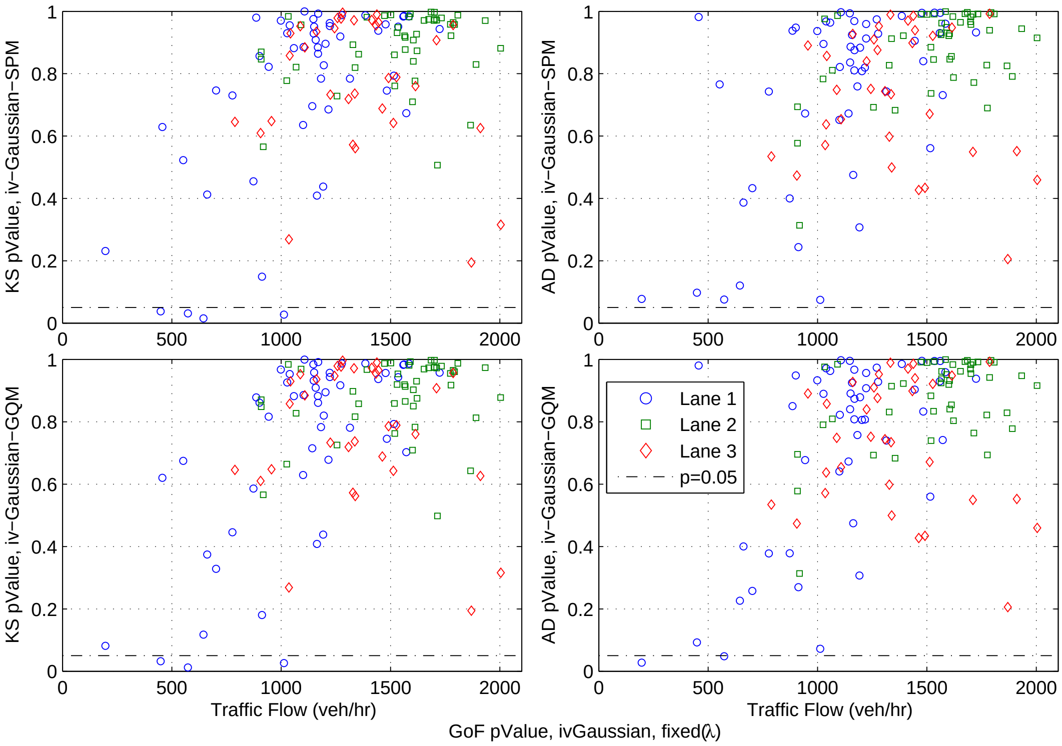

FIGURE 5.13: GoF result (p value) for IVG-SPM and IVG-GQM with pre-determined λ

as shown in Table 5.10, especially for those in lane 2 and 3. The estimated parameters of EMG (μ, σ and τ) or IVG (μ, σ) seem more aggregated within smaller ranges comparing to the results of which the method without the pre-determined λ.

A similar problem caused by the outlier (sample No.189) is found in IVG-mixed models, as discussed in gamma-mixed models (see discussions in section 4.3.2). Apparently, this sample, with its traffic flow as low as 195 veh/hr, is not reasonably represented by the pre-determined λ using the equation in Table 4.5. With this sample removed from the results of lane 1, the variations of μ, σ are greatly reduced (values shown in brackets). More detailed figures can be found in Appendix C (see Figures C.1 - C.4).

Table 5.10: Parameters variation on EMG-mixed and IVG-mixed models with pre-determined λ

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
IVG-SPM						
ϕ	0.29	0.18	0.37	0.11	0.60	0.08
λ	0.43	0.13	0.58	0.12	0.29	0.06
μ	1.87(1.63)	1.73(0.64)	1.31	0.22	1.23	0.13
σ	10.59(10.58)	2.93(2.96)	8.62	1.28	8.69	1.67
IVG-GQM						
ϕ	0.33	0.20	0.45	0.14	0.60	0.08
λ	0.43	0.13	0.58	0.12	0.29	0.06
μ	1.83(1.61)	1.57(0.44)	1.38	0.32	1.24	0.13
σ	10.77(10.75)	2.90(2.93)	8.64	1.29	8.68	1.67
EMG-SPM						
ϕ	0.42	0.23	0.41	0.12	0.62	0.08
λ	0.43	0.12	0.58	0.12	0.29	0.06
μ	0.93	0.12	0.78	0.08	0.73	0.06
σ	0.23	0.07	0.19	0.04	0.17	0.02
τ	1.23	0.96	0.63	0.25	0.57	0.12
EMG-GQM						
ϕ	0.46	0.36	0.56	0.18	0.64	0.08
λ	0.43	0.12	0.58	0.12	0.29	0.06
μ	0.92	0.14	0.80	0.08	0.73	0.06
σ	0.22	0.08	0.20	0.04	0.17	0.02
τ	1.00	0.64	0.76	0.32	0.58	0.13

The estimated mean followers headways (see Table 5.11 and Figure 5.14) are more focused to their average values, especially in lane 2 and 3, across all the four mixed models. In lane 1, the values of mean headway seem more scattered with lower level of traffic flows, and they also become more consistent with heavier traffic.

Greater variations can be found for the IVG-mixed models and this is again caused by the outlier estimates of sample No. 189. Taking away this sample, the data show (in brackets) a better consistency.

There seems to be some decreasing trends in lane 1 for mean headway values, with

TABLE 5.11: Mean Followers Headway(s) with pre-determined λ on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
IVG-						
SPM	1.87(1.63)	1.73(0.64)	1.31	0.22	1.23	0.13
GQM	1.83(1.61)	1.57(0.44)	1.38	0.32	1.24	0.13
EMG-						
SPM	2.16	0.99	1.42	0.25	1.30	0.14
GQM	1.92	0.68	1.56	0.32	1.31	0.14

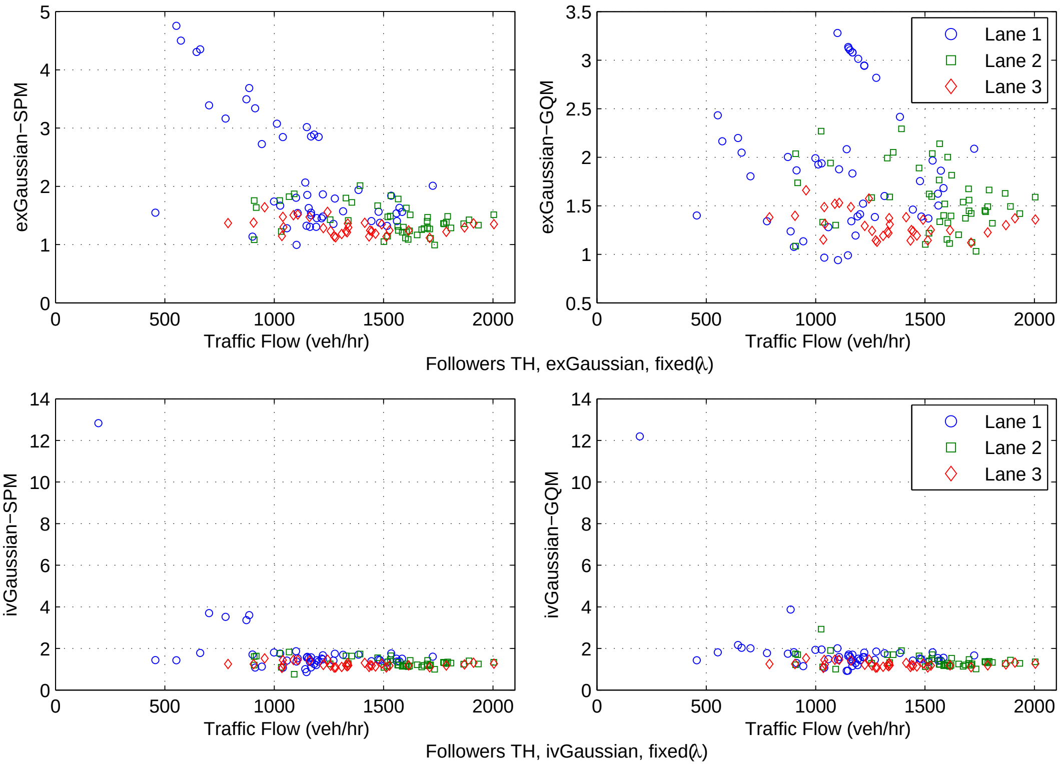

FIGURE 5.14: Mean Followers Headway (s) with pre-determined λ on EMG-mixed and IVG-mixed models. The y-axis show the mean followers headway of each specified model.

increasing traffic flows. However, it is unclear whether it is caused by the changes of traffic flow levels. Two assumptions seem reasonable. The first reason may be that there are bigger variations in queue formulations with lower traffic, which may cause problems for these models to have good estimations to the mean followers headway. It is also possible that, with the increasing but still relatively lower traffic flows, drivers may have increasing tensions that will enforce them to follow preceding vehicles as shorter distances. The increasing tensions will reach a limit when traffic flow reaches a higher level, and hence the mean headway becomes more consistent. Since there is not much data on lower traffic flows for lane 2 and 3, the current available headway data are not sufficient for the investigation of these two assumptions. But this is certainly very attractive to check in future when more headway data becomes available.

Overall, with predetermined λ, all the four mixed models can describe headway samples very well. There are slightly downgrade of fitting performance in lane 1, but the acceptance of the GoF test is still excellent.

5.2.6 Use of Combinations of λ and Other Parameters of Followers Headway Distribution

Similar to the procedures discussed in Section 4.4, some of the combinations of parameters are tested. After a few preliminary tests for each mixed model, some combinations of parameters are identified with good fitting results. For using IVG as followers headway distribution, pre-determining both λ and σ would have better fitting comparing to use combinations of λ and μ. Such better performance, with the combination of λ and σ, are found in both SPM and GQM models. When using EMG model as followers headway distribution, the combination of the three pre-determined parameters, i.e. λ, σ and τ, seems to have the best fitting results compared with other possible combinations of parameters.

Table 5.12 shows the pre-determined values of parameters used in the models fitting stages. For IVG-mixed models, a constant σ, with value of 9.0, is used in fitting all samples. The fitting results show that it is sufficient to use a constant value of parameter σ, which indicates that σ is not sensitive to the traffic conditions, such as flow levels or lanes. However, more careful adjustment according to lanes and traffic flow rates may further improve the fitting results.

TABLE 5.12: Pre-fixed values of parameters on EMG-mixed and IVG-mixed models

Parameter	Lane 1		Lane 2	Lane 3
IVG				
σ	9.00		9.00	9.00
EMG				
Flow (q)	<900	≥900		
σ	0.20	0.20	0.20	0.20
τ	0.80	0.70	0.91	0.56

When using EMG-mixed models, the situation changes from the above. The σ values in EMG model can be pre-determined as a constant value of 0.2. However, it is impossible to find a constant value for parameter τ, which will support models to fit all the data samples. For lane 2 and 3, constant values of 0.91 and 0.56 are used respectively. In lane 1, a further separation have to be made in terms of lower and higher traffic volumes, with a constant value of 0.8 used for flows smaller than 900 veh/hr. For traffic flows greater than 900 veh/hr, the value of 0.7 is found to be more suitable. This distinction of parameters for EMG-mixed models is different when compared to IVG-mixed or gamma-mixed models, as the latter two can both be used with a single constant parameter for all samples. This feature of the EMG model, on the one hand is a disadvantage for models to have a more sensitive parameter comparing those models that can use constant values. However, on the other hand, it may be very valuable to use such a sensitive parameter to find connections between models and variations of external conditions, such as traffic flows, lanes etc. If connections are truly found, it will help greatly in understanding of drivers' car following behaviours. Such work would require large amount of headway data collected in various known traffic situations, but it is certainly worth more investigation with those data available.

The results of GoF tests are summarized in Table 5.13, with more details shown in Figure 5.15 and 5.16. The IVG-mixed models have the best fitting results, with slight better performance comparing to the gamma-mixed models (comparing to Table 4.11). Most of p-values of IVG-mixed models are relative higher values comparing to EMG-mixed and gamma-mixed models. It should be noticed, for lane 1, the fitting results are in fact improved for IVG-mixed models, comparing to the results shown in Table 5.9. This is quite encouraging, as it may indicate the fact that with appropriate constraints of parameter σ, the fitting performance of IVG-models can be more robust.

TABLE 5.13: GoF results with pre-determined λ, σ (and τ) on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
IVG-						
SPM	50(2)	51(1)	47(0)	47(0)	32(1)	32(1)
GQM	50(2)	51(1)	47(0)	47(0)	32(1)	32(1)
EMG-						
SPM	48(4)	51(1)	46(1)	46(1)	32(1)	32(1)
GQM	47(5)	49(3)	47(0)	47(0)	32(1)	32(1)

The numbers with or without brackets show the numbers of successes or failures of GoF test.

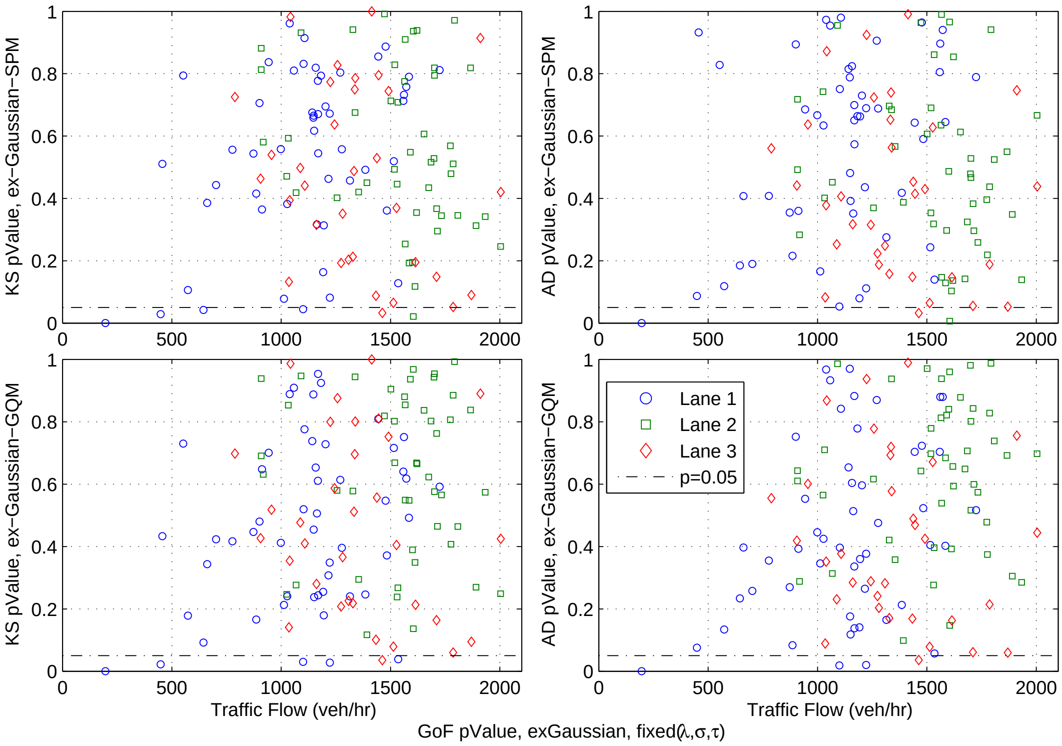

FIGURE 5.15: GoF result (p value) with Pre-determined λ, σ and τ for EMG-SPM and EMG-GQM models

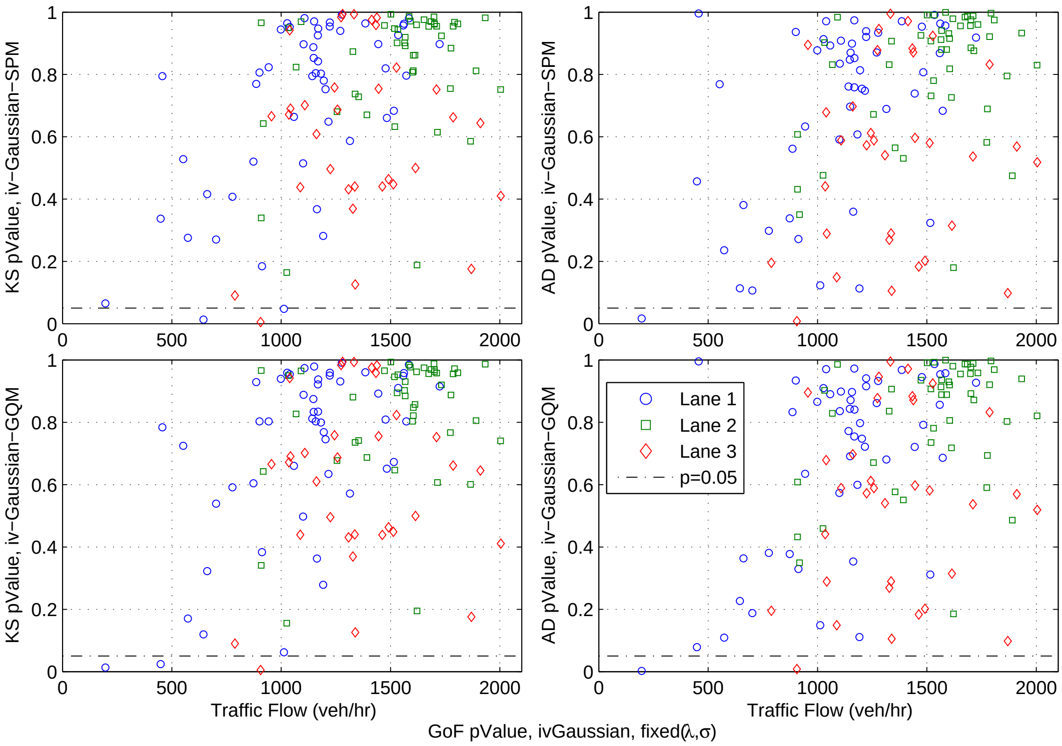

FIGURE 5.16: GoF result (p value) with Pre-determined λ and σ for IVG-SPM and IVG-GQM

The EMG-mixed model has 2-3 more samples, comparing to IVG-mixed models in lane 1, which have failed their KS-test. This shows some but not greatly downgraded in terms of fitting performance. Considering that EMG-mixed models have very good fitting performance shown in the previous sections, this downgrade may be caused by the inaccuracy of pre-determined parameter τ, especially when fitting headway samples of lane 1. In this case, the results may be improved by finding better connections between τ and some traffic situations, such as traffic flows.

When using IVG as followers headway model, both SPM and GQM seem to have very similar fitting performance. While using the EMG model for followers headway distribution, the test results are slightly different between the SPM and GQM models (in lane 1 and 2), however, it is difficult to conclude that one model is better than the other.

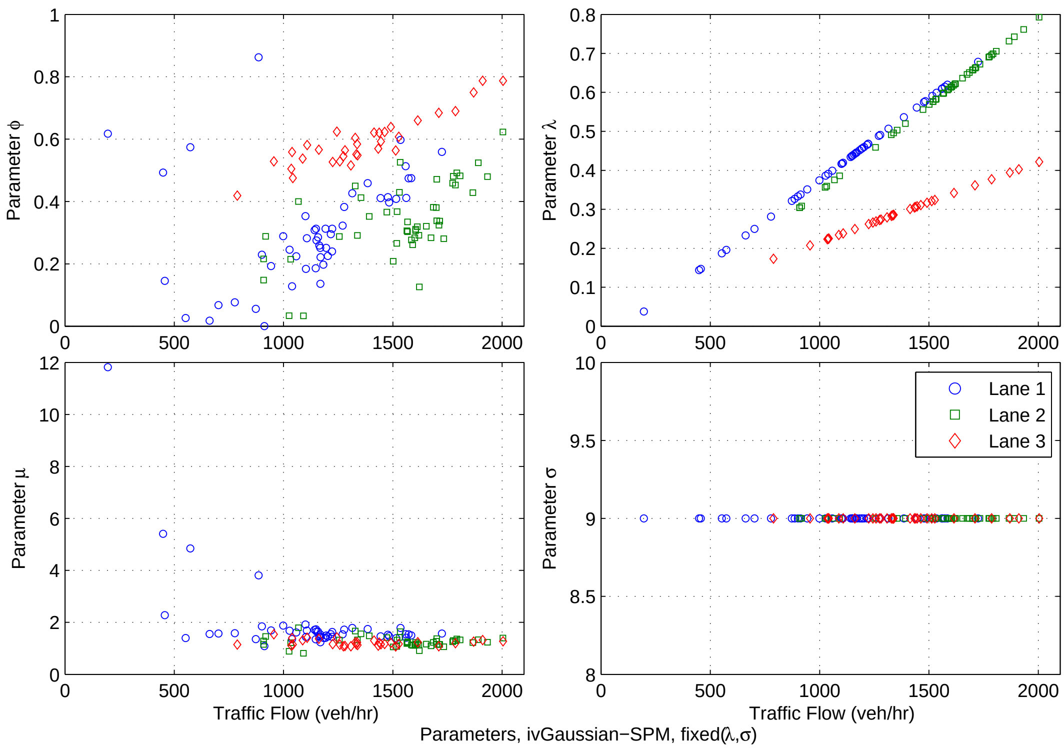

FIGURE 5.17: Estimated parameters with Pre-determined λ and σ for IVG-SPM

TABLE 5.14: Parameters variation with Pre-determined λ, σ (and τ) on EMG-mixed and IVG-mixed models

	Lane 1 Mean	Lane 1 Std	Lane 2 Mean	Lane 2 Std	Lane 3 Mean	Lane 3 Std
IVG-SPM						
ϕ	0.31	0.17	0.34	0.12	0.59	0.08
λ	0.42	0.14	0.58	0.12	0.29	0.06
μ	1.96	1.63	1.25	0.19	1.22	0.12
σ	9.00	0.00	9.00	0.00	9.00	0.00
IVG-GQM						
ϕ	0.38	0.16	0.40	0.14	0.60	0.08
λ	0.42	0.12	0.58	0.12	0.29	0.06
μ	1.77	0.38	1.29	0.20	1.22	0.12
σ	9.00	0.00	9.00	0.00	9.00	0.00
EMG-SPM						
ϕ	0.26	0.14	0.56	0.14	0.63	0.09
λ	0.44	0.12	0.58	0.12	0.29	0.06
μ	0.87	0.05	0.76	0.05	0.75	0.06
σ	0.20	0.00	0.20	0.00	0.20	0.00
τ	0.72	0.04	0.91	0.00	0.56	0.00
EMG-GQM						
ϕ	0.31	0.16	0.64	0.14	0.64	0.09
λ	0.43	0.12	0.58	0.12	0.29	0.06
μ	0.89	0.05	0.78	0.05	0.75	0.06
σ	0.20	0.00	0.20	0.00	0.20	0.00
τ	0.72	0.04	0.91	0.00	0.56	0.00

The results of parameter estimation are shown in Table 5.14 and Figures 5.17 - 5.20, where the pre-determined parameters are marked in grey. With the pre-fixed σ in IVG-mixed models, as well the pre-determined parameter λ, only ϕ and μ are required for estimation. The estimated parameters of μ seem more consistent with the additional constraint on parameter σ. Still, in lane 1, a slight decreasing trend can be observed at relatively lower traffic flows. In IVG-SPM, the noticeable large μ value caused by the sample (No. 189) with very low flow rates makes the estimated values to have large standard deviation. This can be corrected by removing this outlier, the resulted values are shown in brackets for

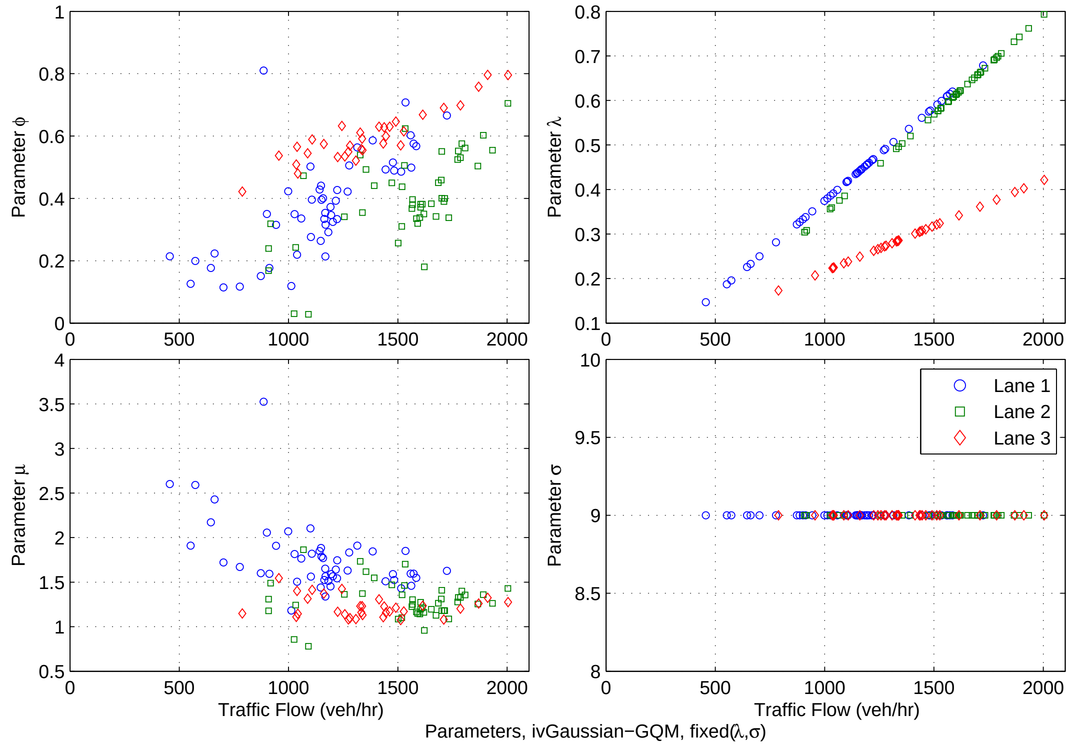

FIGURE 5.18: Estimated parameters with Pre-determined λ and σ for IVG-GQM

IVG-SPM. The same sample did not pass the GoF test stage with IVG-GQM models. The parameter ϕ obtained more reasonable estimates when σ is pre-fixed, in lane 1, and these improvement are clearer shown in the results of IVG-GQM model. Clear increasing trends of estimated ϕ values, with the increased traffic flows, can be observed for both IVG-mixed models.

With EMG-mixed models, the estimated values of parameter μ also become more consistent comparing with the results of early sections. The values of lane 1 seem higher than values of lane 2 and 3, and they don't seem vary much with the changes of traffic flows. The parameter ϕ shows very clear increasing trends with increased traffic flows, for both EMG-mixed models.

In general, the proposed combination of pre-determined parameters, for all the

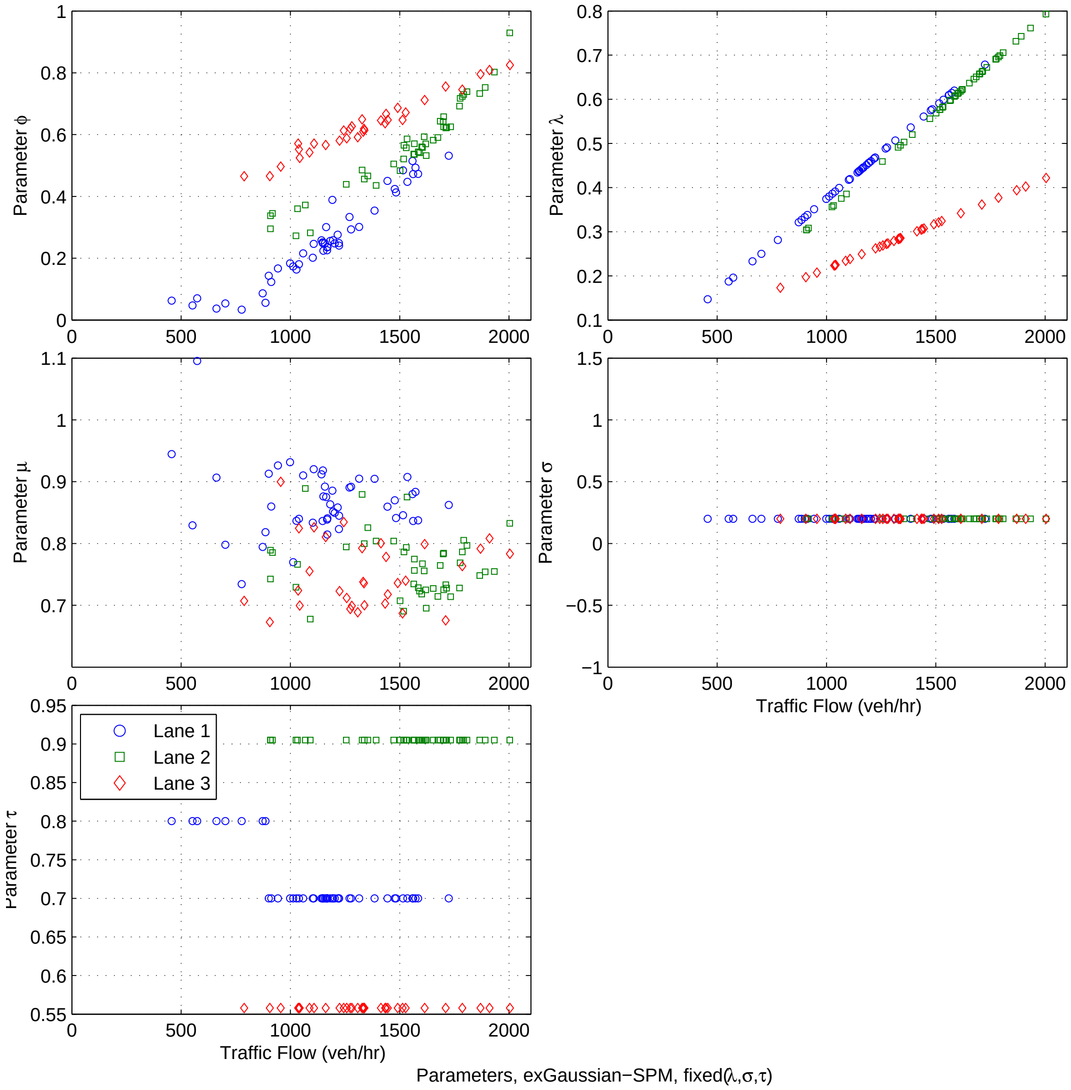

FIGURE 5.19: Estimated parameterswith Pre-determined λ, σ and τ for EMG-SPM

four models, seem to have positive support on parameter estimations. Plus the convincing results shown when fitting with gamma-mixed models using similar method (see Section 4.4.2), these confirm that the proposed methods of simplification are useful in fitting the selected mixed models.

The results of estimated mean followers headway are shown in Table 5.15 and Figure 5.21. Despite to the outlier (sample No. 189), the mean followers headways show very good consistency with all the four models. The slight decreasing trends

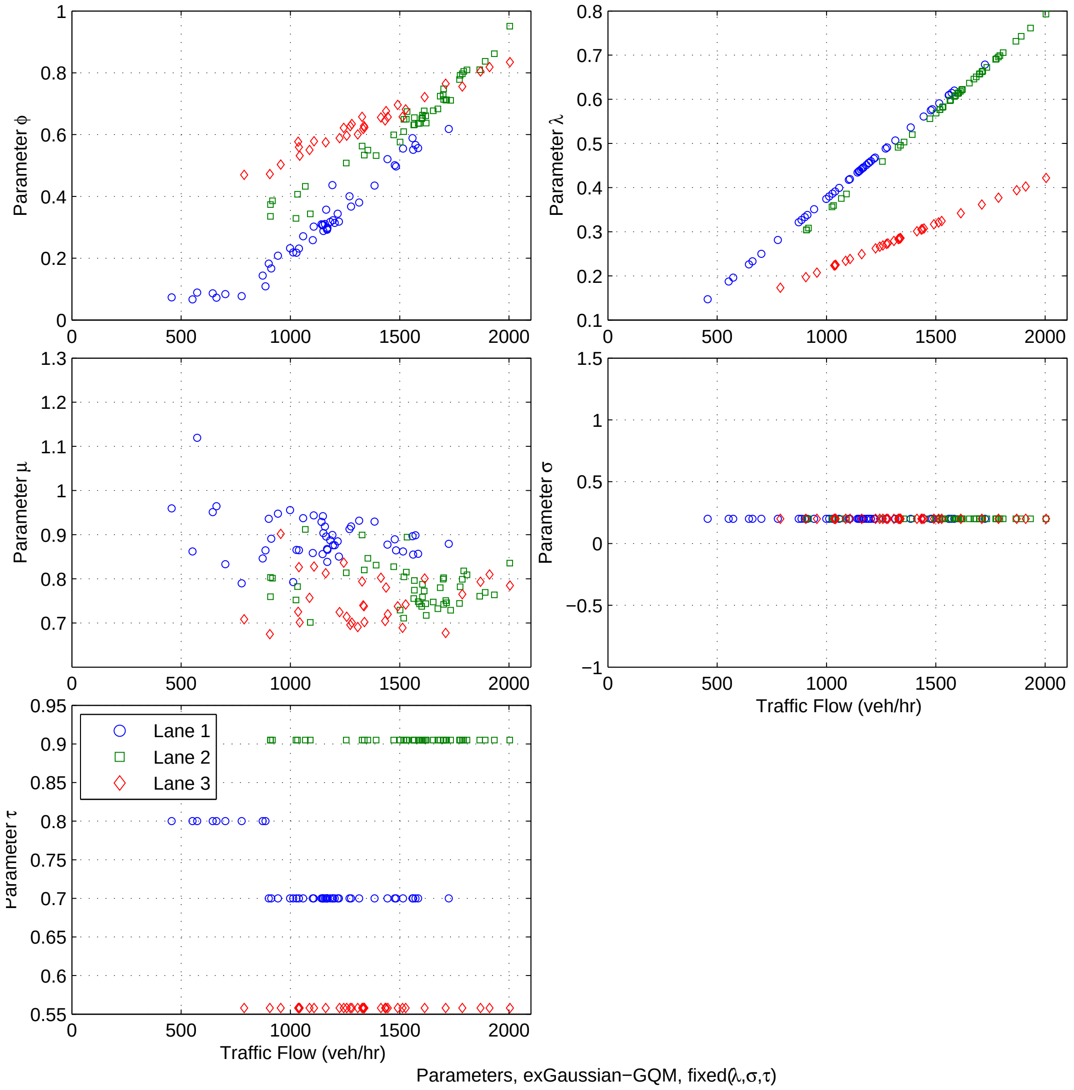

FIGURE 5.20: Estimated parameters with Pre-determined λ, σ and τ for EMG-GQM

can be observed in lane 1, at lower traffic flows. With the higher traffic flows, all the mean followers headways are maintained consistently. The mean headway of lane 1 and 2 are higher than the results of lane 3. Especially with EMG-mixed models, the difference of mean headways can be clearly observed in Figure 5.21.

It is interesting to notice that, for both gamma-mixed and IVG-mixed models, the mean headway in lane 1 seem to be higher than the results in lane 2. However, in EMG models, the opposite result is observed. It is unclear what is the cause

Table 5.15: Mean Followers Headway(s) with pre-determined λ, σ (and τ) on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
IVG-						
SPM	1.96(1.77)	1.63(0.8)	1.25	0.19	1.22	0.12
GQM	1.77	0.38	1.29	0.20	1.22	0.12
EMG-						
SPM	1.58	0.07	1.67	0.05	1.31	0.06
GQM	1.61	0.07	1.69	0.05	1.31	0.06

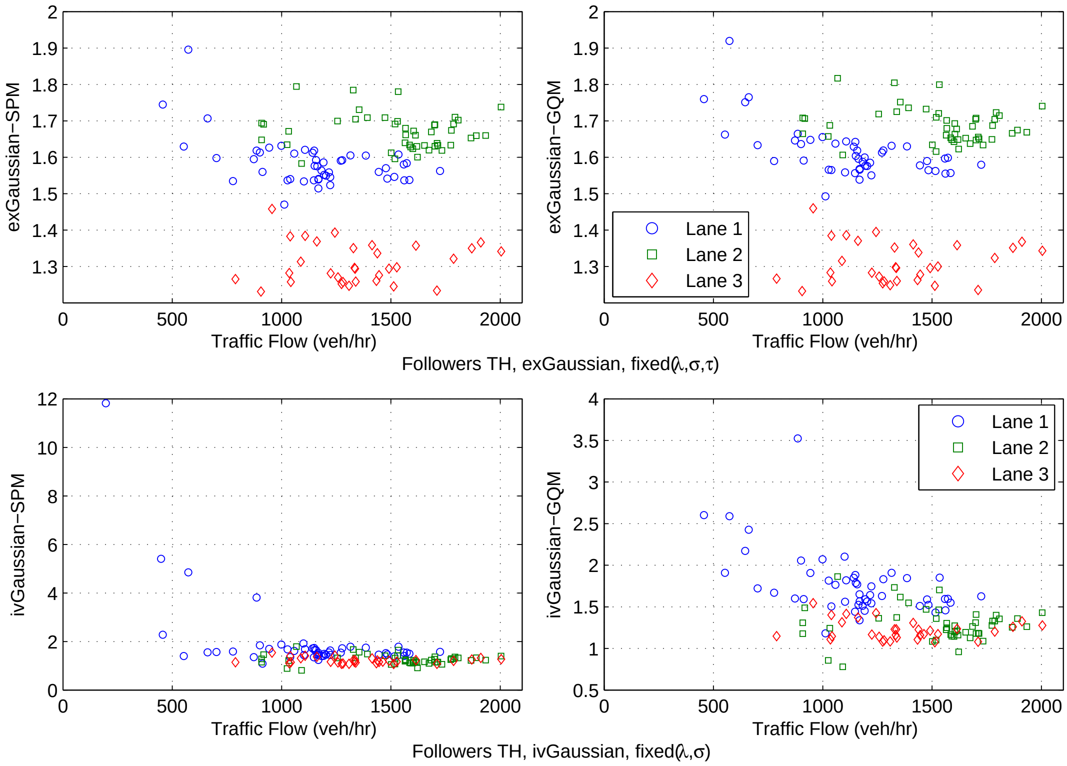

Figure 5.21: Mean Followers Headway (s) with pre-determined λ, σ (and τ) on EMG-mixed and IVG-mixed models. The y-axis show the mean followers headway of each specified model.

of these observations. This may be supposed to have some connection with the pre-determined parameter τ, which has to be separated by traffic flows and lanes in order to maintain better fitting performance, in the use of EMG-mixed models.

Overall, for all the mixed models, including the gamma-mixed models in chapter 4, using the proposed methods which give λ and other parameters reasonable constraint would not greatly affect the fitting performance of those models. The GoF results are slightly downgraded under some traffic flow situations, but the results are still perfectly acceptable. On the positive side, such methods are greatly simplified the estimation of the model by effectively reducing the number of parameters. Also, with these parameters constrained by pre-determination, the proportion factor ϕ receives much more reasonable estimations, which make this parameter more meaningful considering the basic idea of using mixed models, i.e. under higher traffic flow situations, there would be more drivers staying in a queue. Furthermore, the proposed methods result the mean followers headway to more consistency under each model, which positively support the idea that the drivers follow vehicles more consistently when the flow rates reach a higher level, but drivers tensions changes with flow rates when the road is not busy. However, there is not enough evidence to conclude such an assumption, which presumably will demands much deeper investigation. Both Wasielewski (1979) and Branston (1976) supported that followers headway are independent to the changes of traffic flows. There is also an agreement in present study, when traffic flows are roughly more than 1000 veh/h. However, both authors conducted their studies using headway data with minimum flow rates of 900 veh/h, and it is unknown whether there is any similar trend under low traffic flow conditions.

5.2.7 Using Parameter ϕ

As discussed in the previous section, for all the four mixed models, the proportion factor ϕ appears with more apparent trends when apply the simplifying the estimation by using some combinations of pre-determined parameters. Such results make further simplification possible by pre-determining ϕ.

Based on the results of estimating ϕ in the previous section, quadratic equations are obtained by applying polynomial fitting to the estimated values of ϕ, as shown in Table 5.16. These equations are further used to pre-determine ϕ for each of the models, as the final attempts of simplifying the estimation process. With this simplification, only one parameter μ is left to require estimation for each of the four mixed models.

TABLE 5.16: Estimation of ϕ as function of traffic flow q

	IVG			EMG		
	Lane 1	Lane 2	Lane 3	Lane 1	Lane 2	Lane 3
SPM						
a	3.1	3.51	0.93	3.7	4.7	0.05
b	-0.14	-1.77	0.25	-0.84	-1.99	1.06
c	0.01	0.44	0.35	0.11	0.54	0.22
GQM						
a	3.1	3.09	0.93	3.42	2.5	0.05
b	-0.14	-1.30	0.25	-0.26	-0.12	1.07
c	0.01	0.34	0.35	0.03	0.22	0.23

a, b and c are parameters of function $\phi(q)$, where $\phi(q) = a * q^2 + b * q + c$, and q is sample traffic flow.

All the fitting times (Table 5.17) are reduced to around 0.12 seconds in average, with standard deviation of 0.02. This is on the same level of gamma-mixed models, also with ϕ being pre-determined. As expected, there are not any improvement on speed of GoF test for IVG-SPM and EMG-SPM models, as numerical integrations involved in their CDFs are inevitable.

TABLE 5.17: Comparison of estimation and GoF time on EMG-SPM, EMG-GQM, IVG-SPM and IVG-GQM models

	Fitting Time(s)		KS test time(ms)		AD test time(ms)	
	Mean	Std	Mean	Std	Mean	Std
IVG-						
SPM	0.12	0.02	343	124	360	159
GQM	0.11	0.02	7	11	2	1
EMG-						
SPM	0.12	0.02	344	117	362	161
GQM	0.12	0.01	6	7	3	1

TABLE 5.18: GoF results with pre-determined ϕ, λ, σ (and τ) on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	KS	AD	KS	AD	KS	AD
IVG-						
SPM	48(4)	50(2)	44(3)	45(2)	27(6)	29(4)
GQM	48(4)	49(3)	42(5)	44(3)	25(8)	28(5)
EMG-						
SPM	49(3)	50(2)	47(0)	46(1)	31(2)	28(5)
GQM	46(6)	48(4)	47(0)	47(0)	31(2)	29(4)

The numbers with or without brackets show the numbers of successes or failures of GoF test.

Generally, the GoF tests show well accepted results for all the 4 models across all samples, although they are downgraded comparing to previous methods. The IVG-mixed models have similar fitting performance comparing to gamma-mixed models. The EMG-SPM shows slightly better fitting performance comparing to all the rest of models including gamma-mixed models.

The fitting performance of IVG-mixed model are downgraded slightly when compared to the results of the previous methods, which did not use the pre-determined

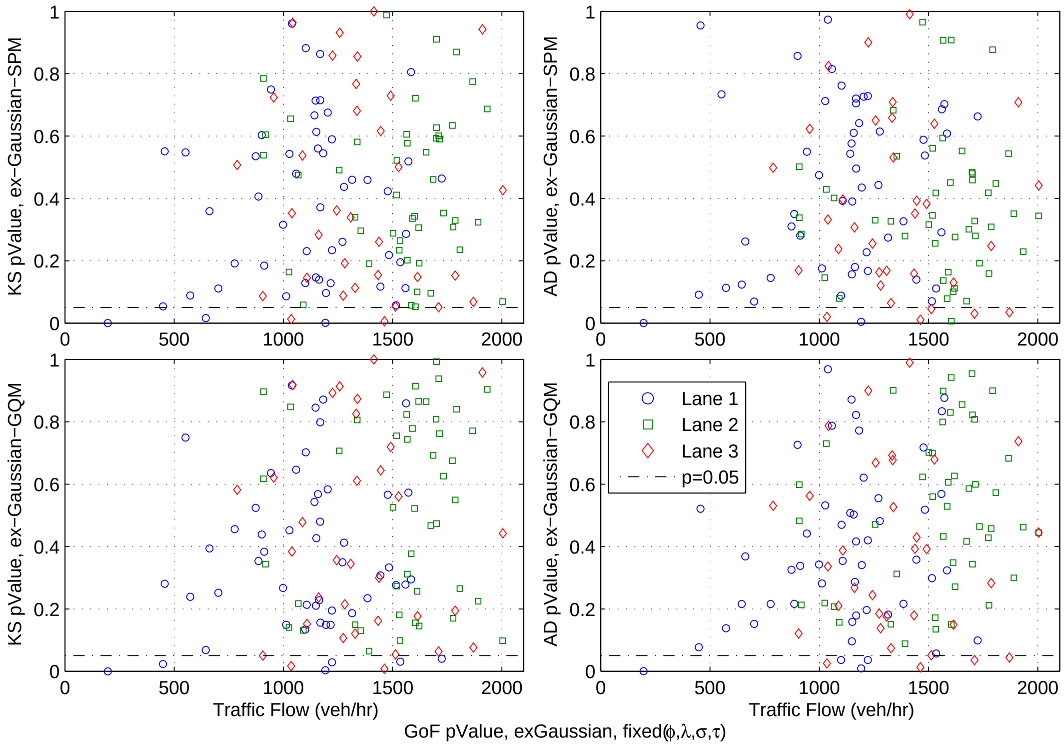

FIGURE 5.22: GoF result (p value) with Pre-determined ϕ, λ, σ and τ for EMG-SPM and EMG-GQM models

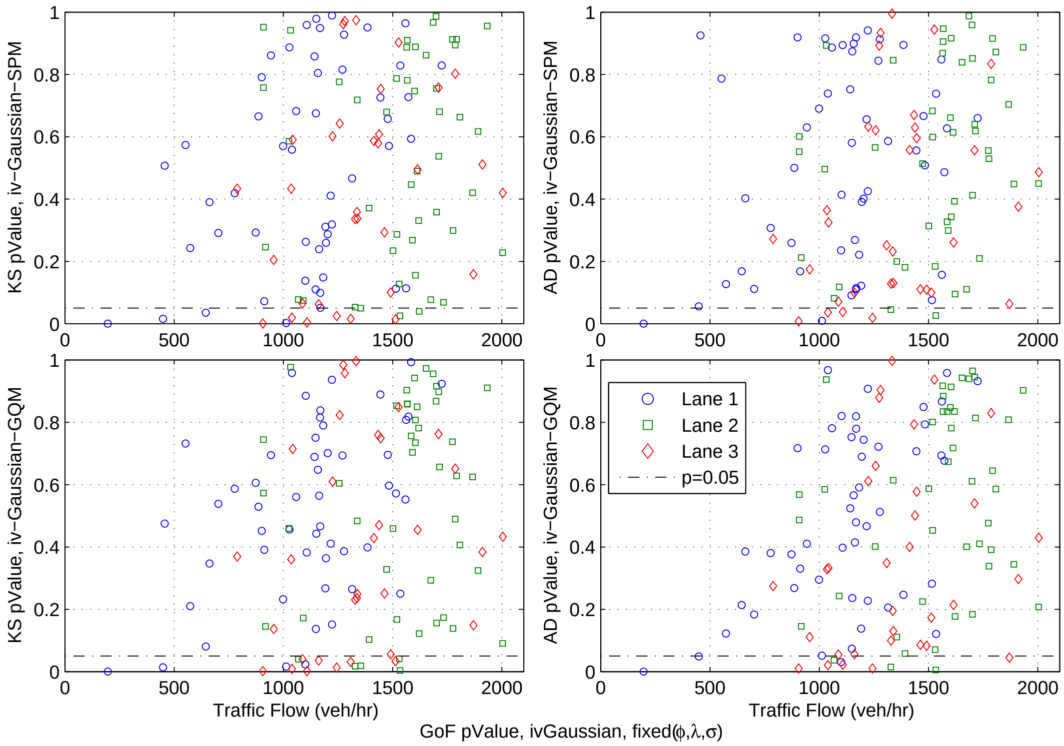

FIGURE 5.23: GoF result (p value) with Pre-determined ϕ, λ and σ for IVG-SPM and IVG-GQM

ϕ for simplification. In contrast, with the EMG-mixed models, the fitting results are not affected much. Especially in lane 3, where more samples failed GoF tests when using gamma- and IVG-mixed models. This comparison shows a better stability on fitting performance when using EMG-mixed models.

Only the parameter μ for IVG-mixed and EMG-mixed models is needed to estimate in this stage (see Table 5.19). Compared with the previous section, the parameter of μ has not varied much in lane 2 and 3, and they seem gathered more closely to the mean values, with smaller variance. In lane 1, the values of μ of EMG-models have not changed much either. The standard deviations of values in IVG-mixed models has been greatly reduced, which indicate that these estimated μ values also become more consistent. Detailed figures can be found in Appendix C (Figures C.5 - C.8).

The mean followers headways do not differ much from the estimations of previous section. The smaller standard deviation especially in IVG-models show more consistency on mean followers headway. The values don't differentiate between SPM and GQM models, and they don't seem to have any trend to the changes of traffic flow, in lane 2 and 3. At the lower traffic flows in lane 1, the decreasing trend of μ can still be observed, which shows that the parameter μ is not affected much by pre-determining parameter ϕ. It seems more likely that this trend is correlated to the traffic flow levels, rather than a random effect. If more headway samples are available in lane 2 and 3, with the similar low level of traffic flows, this trend may be revealed more explicitly.

TABLE 5.19: Parameters variation with Pre-determined ϕ, λ, σ (and τ) on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
IVG-SPM						
ϕ	0.30	0.14	0.34	0.08	0.59	0.08
λ	0.44	0.12	0.58	0.13	0.30	0.06
μ	1.64	0.11	1.25	0.07	1.20	0.09
σ	9.00	0.00	9.00	0.00	9.00	0.00
IVG-GQM						
ϕ	0.30	0.15	0.38	0.09	0.59	0.08
λ	0.43	0.12	0.59	0.13	0.30	0.06
μ	1.57	0.09	1.24	0.05	1.19	0.08
σ	9.00	0.00	9.00	0.00	9.00	0.00
EMG-SPM						
ϕ	0.24	0.12	0.57	0.13	0.63	0.09
λ	0.43	0.13	0.58	0.12	0.29	0.06
μ	0.88	0.07	0.77	0.04	0.75	0.06
σ	0.20	0.00	0.20	0.00	0.20	0.00
τ	0.72	0.04	0.91	0.00	0.56	0.00
EMG-GQM						
ϕ	0.30	0.14	0.64	0.14	0.64	0.09
λ	0.42	0.12	0.58	0.12	0.29	0.06
μ	0.90	0.05	0.78	0.04	0.75	0.06
σ	0.20	0.00	0.20	0.00	0.20	0.00
τ	0.72	0.04	0.91	0.00	0.56	0.00

Looking at all the gamma-mixed and IVG-mixed models, the values in lane 1 is generally larger than in lane 2, of which the values are slightly larger than in lane 3. The values in lane 2 and 3 are much closer (around 1.2s) and they have similar values for both gamma-mixed and IVG-mixed models. The EMG-mixed models seem to estimate the mean followers headway more closely (around 1.6s) in lane 1 and lane2, but smaller (1.3s) in lane 3. The values in lane 2 seem again higher than values in lane 1, with higher traffic flows (shown in Figure 5.24).

Figure 5.25 shows the visual fitting result for six selected headway samples. The estimated PDF curves show good fitting results visually for all the four models.

TABLE 5.20: Mean Followers Headway(s) with pre-determined ϕ, λ, σ (and τ) on EMG-mixed and IVG-mixed models

	Lane 1		Lane 2		Lane 3	
	Mean	Std	Mean	Std	Mean	Std
IVG-						
SPM	1.64	0.11	1.25	0.07	1.20	0.09
GQM	1.57	0.09	1.24	0.05	1.19	0.08
EMG-						
SPM	1.60	0.10	1.67	0.04	1.31	0.06
GQM	1.62	0.07	1.69	0.04	1.31	0.06

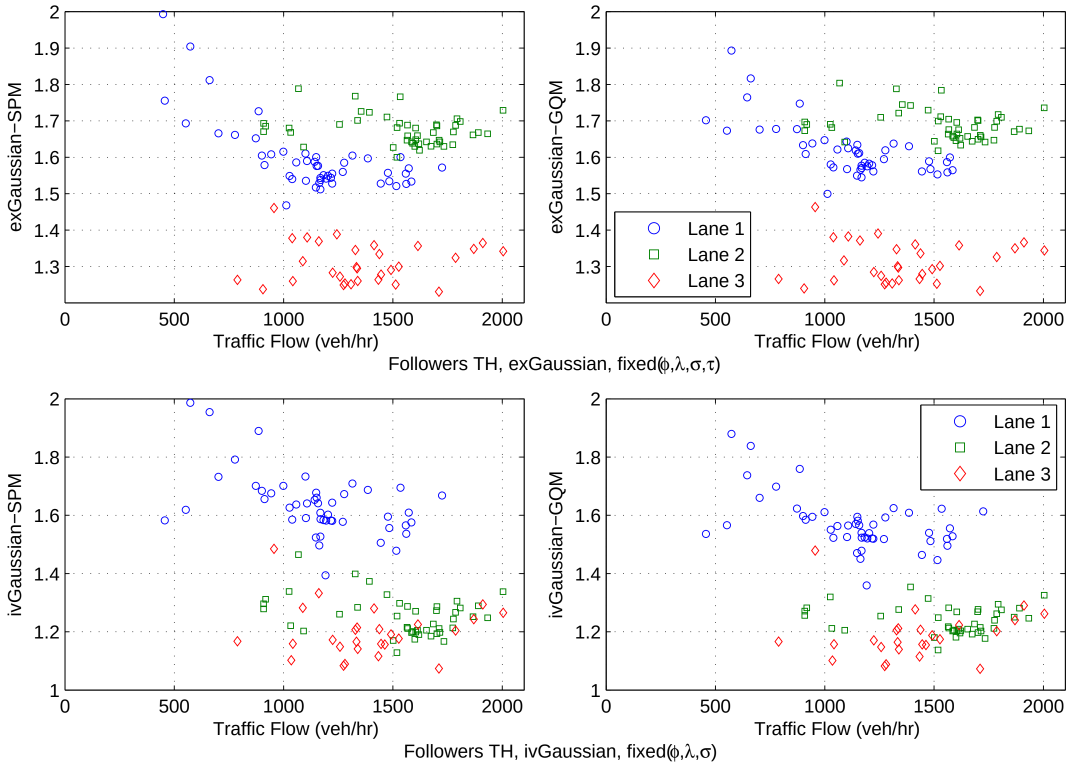

FIGURE 5.24: Mean Followers Headway (s) with pre-determined ϕ, λ, σ (and τ) on EMG-mixed and IVG-mixed models. The y-axis show the mean followers headway of each specified model.

Overall, this final attempt on simplification of parameters only estimated one parameter (μ), with all the rest of the parameters pre-determined. The good acceptance of the GoF result shows that all mixed models can be simplified greatly, without losing much of their fitting performance. It is worth highlighting that the

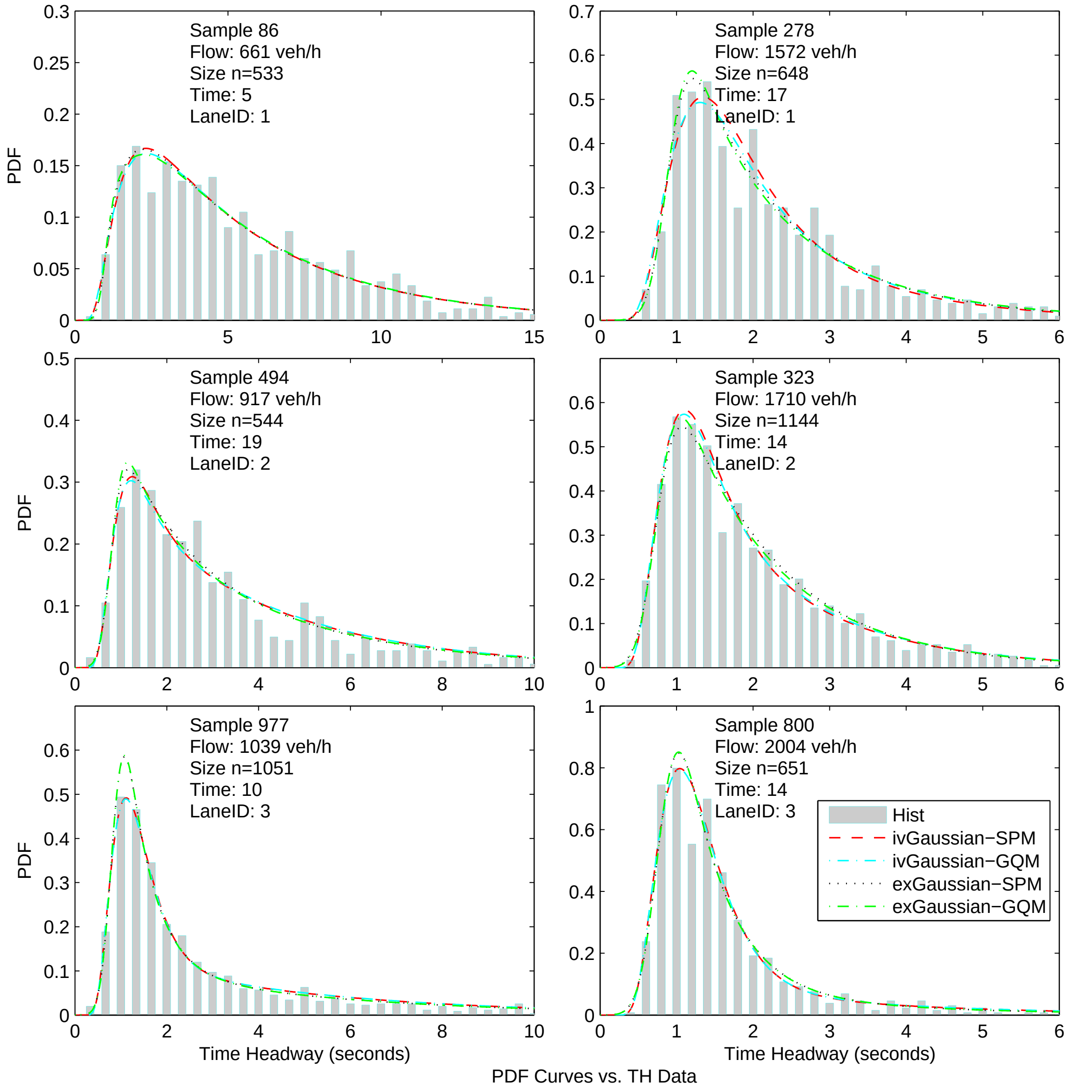

FIGURE 5.25: PDF curves with pre-determined ϕ, λ, σ (and τ) on EMG-mixed and IVG-mixed models

two EMG-mixed models, as they have the best fitting performance after simplification, compared with the gamma-mixed and IVG-mixed models.

These results are very positive and encouraging, as they are strong evidence to support that each of the parameters in mixed models is worth further investigation, as they may have strong connections to the corresponding traffic conditions or variations of external traffic environment. These potential connections may be good hints for further understanding the drivers' car following behaviours.

5.3 Review and Discussion

Two response time (RT) models, i.e. the *ex-Gaussian*(EMG) and *iv-Gaussian*(IVG) distribution models are, for the first time, used in headway analysis. The results of both models are acceptable on fitting the headway data of lane 1 and 2. These results are comparable with the *shifted-lognormal* distribution, as they have similar fitting performance for the first two lanes of the motorway. However, none of these models shows acceptable fitting results for lane 3. Nevertheless, these results show a good evidence that, in general, RT distributions have very good potentials to be good models to describe headway data.

Both EMG and IVG distribution models are then applied as the followers headway distribution in the SPM and GQM mixed models. Using the numerical MLE methods, all the four mixed models show reasonable estimation time of under 1 second. However, since the SPM models require numerical integration involved in CDF, they lead to slower speed on the GoF test. All the four models are well fitted to the headway samples, which are slightly better than the gamma-mixed and lognormal-mixed models tested in chapter 4. Still, the 4- or 5-parameter models are too flexible and some of the estimations, for headway data in lane 1, seem unrealistic.

Overall, it is difficult to distinguish firstly the types of models between SPM and GQM as they have similar fitting performance, when using any of the followers headway distributions in this study. The downside of the SPM model is that it requires numerical integrations in computing CDFs, and this is often difficult or maybe impossible to be improved by achieving an analytical form of the solution. This may lead to a high computational cost of the GoF test, as Ha et al. (2012) reported for the gamma-SPM model. The SPM models also require the Laplace transform of the PDF of the followers headway distribution, which may not have an explicit function for some models, such as *lognormal* distribution. The GQM models generally have acceptable computational costs if a closed form of Laplace transform exists for the PDF of the followers headway distribution. Otherwise, the estimation time would be unbearable. Considering these properties, GQM may be easier to implement compared with the SPM models, with comparable fitting performance.

All the *gamma*, IVG and EMG models show similar fitting performance when they are used as followers headway distributions in mixed models. The IVG-SPM and EMG-SPM have slower performance while performing the GoF tests. The IVG-GQM is slower with unstable estimation times while fitting to headway data when all four parameters are required for estimation. the EMG-GQM and both gamma-mixed models have stable and acceptable computational costs.

It is hard to say which of these models can be recommended further, considering the close performances on fitting and computational costs. It may require more headway data from various location and known environments, to distinguish that any of this models would be more suitable for use. Considering the EMG-mixed and IVG-mixed models are the first time ever using in the headway data analysis, the at least equivalent performances to gamma-mixed models make applications of these newly proposed models very promising.

With pre-determination of some parameters in those models, the GoF results show very good fitting performance when using fixed λ, or a combination of λ and σ in IVG-models, or combination of λ with σ and τ in EMG-models. These results are comparable with similar performance while using similar constraints in the gamma-mixed models. The proportion factor ϕ is used as an additional constraint, and fitting results are slightly degraded, but are still well accepted in general, like those in gamma-mixed models. In such cases, the EMG-SPM shows the best fitting results compared with the other models. The good fitting results show the success of the attempts of the simplification methods, and they may indicate strong connections between parameters and traffic situations, which are potentially very valuable for understanding drivers' following behaviour.

Chapter 6

Discussion and Future Work

6.1 Final Discussion

Research work on vehicle time-headway often draws attention in many respects, from the fundamental understanding of the traffic formulation to traffic simulation. Statistical analysis is a powerful tool to help researchers and traffic engineers with better understanding of the empirical traffic data, hence, to help on providing improved or new designs of solutions to realistic problems.

This thesis mainly looked at statistical models as an analytical tool of knowing headway data, on motorways in particular. It firstly introduced a framework developed using the MATLAB Statistical Toolbox. The proposed framework provides data structures and implementable solutions that can quickly test or repeat existing statistical models, or making rapid trials for newly proposed models. The framework is not yet ready to use under practical situations as it sacrifices some of the computation performance to provide more generalized solutions to suit more distribution models. But it is still valuable for quick implementation in preliminary research or to help newcomers to rapidly test some ideas or knowledge about headway distribution models.

The second part of the thesis focuses on some commonly used mixed models, i.e. the SPM and GQM models. The common *gamma* and *lognormal* distributions, as followers headway models, are reviewed. The implementation results are similar to those reported by previous researchers. In addition, the lognormal-SPM, for first time, has been tested and reported here, by using an approximate solution for the Laplace transform of the *lognormal* distribution, which has not been used in this context before. However, this is only used for demonstration, as the numerical approximations involved have unavoidable computational cost compared with other models that have explicit solutions in Laplace transform. The implementation of the lognormal-SPM again gives well-accepted fitting results for empirical headway data, and they were very similar to the lognormal-GQM, which has been widely studied in the past.

Following the general implementations of the above models, some parameters have been investigated further, mainly the arrival rates λ and followers headways. Some researchers had investigations on the parameter λ, and their results show good potentials of pre-determining λ, using the known sample flow rate. Similar studies are replicated here, as a simplification in fitting stages. The results of the GoF tests shown are promising as most of data are well accepted with this method.

Followers headways are mainly determined by drivers' perception when they are closely following their preceding vehicles, and it could relate to safety distance, reaction and response time, etc. This gap between two consecutive vehicles may vary under the changes of traffic situation or environment, but it should not deviate dramatically. This may suggest that the follower headway distributions should be rather consistent, with slight variations responding to various environments. A few researchers tried to fix followers headway in mixed models to simplify their estimation process. Some reported very good results when they used constant followers headways. Similar procedures are also attempted in this study, for the gamma-mixed models. However, the result did not give ideal performance as shown by others. The results of lane 1 and 2 are very encouraging as they are close to previous findings. However, more samples are rejected by the GoF test for samples of lane 3, and this may be due to the increasing complexity of traffic of this lane. Alternative methods were investigated, and the attempts at fixing parameter α in the *gamma* followers headway model gave some promising results, combined with pre-determined λ, the two pre-set parameters seem successfully restricted the two 4-parameter models, and provide more sensible solutions. In particular, the estimation of ϕ becomes more reasonable and it shows clear connections with the sample traffic flow. This fixes some of the problematic estimations for ϕ, when 4 parameters are all freely estimated using numerical MLE methods.

Followers headway is analogous to some process in the human recognition and response studies. Two RT models, which are *ex-Gaussian* and *inverse Gaussian* distributions, are investigated in the chapter 5. As single models, both give good fitting results to some of the samples, especially in lane 1 and 2. But all failed in fitting any of those samples in the third lane. These two models offer outstanding and comparable performance compared with the *lognormal* distribution, which is commonly agreed as the most acceptable single model. However, they are still very limited when generally describe the headway data.

For the mixed models, both above RT models are investigated as followers headway distributions, and their fittings to headway data are very well accepted by GoF tests, with all samples having passed those tests. This shows that using RT models is potentially a very good way to investigate the followers headway process, especially when they are equipped with mixed models. Similar to chapter 4 in the gamma-mixed models, some simplification attempts were also investigated for these new mixed models. Again, by limiting parameter λ, hence combined with more parameters in EMG and IVG models, the GoF tests have shown comparable fitting performance to those methods used with gamma-mixed models.

The above concludes that the tested mixed models have capability to describe the headway data, with very good fitting results. However, 4 or 5 parameter models are overly flexible and the optimized solution may not be the best estimation to match the real traffic situation. Using some sensible constraints in the estimation process is necessary, and these will narrow the valid range of values and promote more reasonable estimations. This study would not be able to answer which model would be the best headway data analysis, due to lack of understanding of the process of how drivers would follow their preceding vehicle. But the behaviours of some parameter in those models would help on understanding such process. The question is not which model gives the best fit, but which models would include parameters based on physical features that make a better understanding for the following process.

This study introduces followers headway distributions which are alternatives in terms of providing good fit to headway data using mixed models. Study shows that the RT models are suitable as followers distributions, with at least equivalent performance comparing to gamma-mixed or lognormal-mixed models. This certainly will give researchers more options to choose, and more comparisons among more models.

The followers headway distribution is critical in mixed models; however, variations among different models have been revealed by the study results. That is, although it is possible to use identical equations to describe the non-followers arrival rates λ, the proportion parameter ϕ are not uniform across different models. This is certainly not satisfactory, as in reality there will only be one proportion value per sample, and this proportion should not be dependent on the followers headway distribution.

A similar problem is also expressed by the estimation of parameters of the followers headway model. This is easily found by comparing the mean followers headway. As Wasielewski (1979) and Branston (1976) assumed, the mean followers headway should be constant. It is natural to expect that this mean value would vary, within a small range of values, under certain conditions. For instance, the mean headway may have changes depending on the travel time, i.e. it may vary between morning and afternoon traffic, due to changes in tiredness, motivation etc., which may differ at different periods of the day. The mean value may also vary with the traffic volume, as busy traffic may cause anxiety or frustration to a large group of drivers. These changes in mean values are expected, and it is interesting to find the physical reason. However, changes in mean values are not desirable, when they are dependent on the changes of models. This dependence on models cannot be observed if there is no alternative models available. With more models available to describe followers headway, the estimated mean values can now be compared, and modified by adjusting some constraints to parameter estimations.

A possible solution to this may be to perform a large numbers of tests across models using more headway data. The results may show better and more uniform estimations of parameter ϕ. If both ϕ and λ can be reasonably determined, then the followers mean headway is more reasonably restricted by these two parameters. This may force the mean followers headway to converge to more constant values across models. If this can happen, it will be a very good reason to believe that the mean headway is much closer to the true values in reality. Hence, all parameters involve in mixed models can be well interpreted with connection to any potential factors that may interfere with traffic flows.

6.2 Future Work

Striving towards a more sophisticated framework in MATLAB will be very valuable. This should mainly be focused on the following aspects:

1. Providing implementations of data sampling methods and fitting methods that have been used in previous studies;

2. Providing more comprehensive functionalities with better GUI, that will make implementation of models simpler.

3. Providing more sophisticated report generators

This work will greatly assist a comparison of headway studies to benefit its users.

Testing the studied models with a large amount of headway data is desirable, especially data from various sites and traffic conditions. This will help in further understanding those parameters in suggested models, hence finding more clear connections between them and potential factors such as traffic flow, hours of the day etc.

In particular, more headway data with lower traffic flows (presumably under 1000 veh/hr) are desirable, as they will provide more information on supporting or rejecting the independence of followers headways to traffic flows.

More investigations with IVG- and EMG-mixed models are important, for their mathematical properties, moment generating functions, random number generations, and so on. These further studies will make the models more applicable for use in reality. More sophisticated fitting methods should be sought, for more consistent and effective parameter estimations. Methods such as Monte Carlo studies should be consulted to properly measure the performance of fitting methods.

6.3 Future Publications

Following this study, some findings are planned to publish in near future. Two main aspects of this work can be prepared with materials of this thesis. These include:

1. Following the work of Asmussen et al. (2014), i.e. the approximation of Laplace transform of *lognormal* distribution, the lognormal-SPM mixed model can now be used for headway analysis. The implementation of this model and its comparison with lognormal-GQM are planned to publish as a journal paper. This paper will include the data preparation and description, the approximation method of Laplace transform for *lognormal* distribution, descriptions of these two lognormal-mixed model and their implementations. Then the results of implementation and comparison will be focused. Discussions of the advantage and deficiencies of these two models will also be included.

2. The second paper will focus on the use of RT models in the study of headway analysis. This paper will briefly discuss the use of the existing single and mixed models in headway analysis. Then the EMG and IVG distributions as single models and as followers headway distributions in mixed models will be introduced. Their implementation methods and test results will be provided. This is followed by comparison of these models and discussion of their fitting performances. Parameter simplification of the mixed models will be discussed, and the combination of the proposed pre-determined parameters and their test results will be given.

The improvement of the results and advantages of using such implementation will also be included.

Appendix A

Time-Headway Data Samples

Table A.1: Statistical Properties of Time-Headway Samples.

ID	L	Size	D	Hour	Flow	Speed	mTH	sTH	CV	Sk.	Kur.	pLGV
63	1	507	1	8-8h	1725	45.72	2.09	1.18	0.57	1.61	6.30	0.15
69	1	801	1	11-11h	1158	55.32	3.11	2.13	0.68	1.71	6.45	0.43
70	1	2401	1	12-14h	1223	54.85	2.94	2.00	0.68	1.90	8.54	0.42
71	1	549	1	14-14h	1271	54.00	2.83	2.02	0.71	2.07	8.91	0.42
76	1	501	1	16-16h	1483	47.38	2.43	1.63	0.67	1.83	7.06	0.20
79	1	667	1	17-18h	1444	47.46	2.49	1.89	0.76	2.73	15.03	0.14
80	1	526	1	18-18h	1192	47.46	3.02	2.43	0.80	1.90	6.91	0.19
81	1	872	1	19-20h	873	57.76	4.12	2.79	0.68	1.51	6.17	0.33
86	1	533	2	5-6h	662	57.12	5.44	3.58	0.66	1.19	4.29	0.59
98	1	2412	2	10-12h	1195	54.94	3.01	2.11	0.70	1.93	8.95	0.43
99	1	1532	2	13-14h	1222	54.94	2.95	2.02	0.68	1.71	7.01	0.41
103	1	1251	2	15-16h	1477	51.80	2.44	1.62	0.66	1.92	7.58	0.32
112	1	551	2	19-20h	901	56.24	4.00	2.83	0.71	1.52	5.44	0.34
120	1	573	3	5-6h	573	58.70	6.28	4.48	0.71	1.97	10.18	0.52
128	1	1592	3	10-11h	1216	55.00	2.96	2.08	0.70	1.84	7.45	0.40
139	1	501	3	14-14h	1559	45.15	2.31	1.67	0.72	2.67	12.71	0.22
151	1	751	3	17-17h	1584	46.49	2.27	1.55	0.68	2.24	9.49	0.12
162	1	591	4	5-6h	449	61.45	8.02	6.39	0.80	2.00	7.96	0.34
164	1	737	4	7-8h	702	61.45	5.13	3.62	0.71	1.72	7.68	0.23
167	1	693	4	8-9h	1012	59.99	3.56	2.51	0.71	1.49	5.81	0.14
168	1	771	4	9-10h	1168	58.23	3.08	2.09	0.68	1.44	5.16	0.14
169	1	1051	4	10-11h	1203	56.58	2.99	1.99	0.66	1.52	5.67	0.13
170	1	801	4	11-12h	1515	52.24	2.38	1.70	0.72	1.91	6.81	0.11
173	1	575	4	13-14h	1163	58.38	3.09	2.20	0.71	1.49	5.19	0.12

L - LaneID; D-Day; mTH - mean Headway; sTH - st.dev. of Headway.

Continued...

ID	L	Size	D	Hour	Flow	Speed	mTH	sTH	CV	Sk.	Kur.	pLGV
175	1	981	4	14-15h	1182	59.06	3.05	2.02	0.66	1.39	4.83	0.12
177	1	520	4	17-17h	1039	58.24	3.46	2.43	0.70	1.67	6.57	0.11
178	1	1000	4	17-18h	943	58.24	3.82	2.75	0.72	2.00	9.01	0.10
181	1	538	4	19-20h	886	61.11	4.06	2.63	0.65	1.40	5.26	0.11
186	1	913	4	21-23h	552	62.95	6.52	5.23	0.80	1.96	8.52	0.11
187	1	601	4	23-0h	457	62.95	7.88	6.61	0.84	1.74	6.52	0.10
189	1	685	5	1-4h	196	62.95	18.38	16.98	0.92	1.73	6.56	0.21
200	1	701	5	9-10h	999	56.98	3.60	2.52	0.70	1.68	6.39	0.11
203	1	2037	5	11-12h	1149	56.14	3.13	2.13	0.68	1.91	8.74	0.11
204	1	1951	5	12-14h	1102	57.08	3.27	2.30	0.71	1.82	7.64	0.14
205	1	551	5	14-15h	1142	55.46	3.15	2.21	0.70	1.90	7.50	0.17
206	1	2251	5	15-17h	1147	57.14	3.14	2.27	0.72	1.82	7.32	0.15
208	1	1060	5	17-18h	1107	58.93	3.25	2.29	0.70	1.96	9.02	0.15
209	1	617	5	18-18h	1027	58.93	3.50	2.50	0.71	1.71	6.65	0.14
210	1	1309	5	18-20h	912	58.93	3.95	2.71	0.69	1.50	5.70	0.16
211	1	558	5	20-21h	777	60.67	4.63	3.17	0.68	1.64	7.13	0.17
234	1	1101	6	8-9h	1561	46.39	2.31	1.60	0.69	2.19	9.39	0.18
236	1	1250	6	10-11h	1168	55.86	3.08	2.13	0.69	1.74	6.80	0.41
238	1	5051	6	11-15h	1169	55.06	3.08	2.12	0.69	1.66	6.38	0.43
239	1	513	6	15-16h	1277	54.67	2.82	1.83	0.65	1.57	5.48	0.35
241	1	551	6	17-18h	1315	47.24	2.74	1.71	0.62	1.56	5.82	0.24
257	1	801	7	7-7h	1535	45.22	2.35	1.47	0.63	2.28	10.99	0.18
266	1	508	7	10-10h	1058	54.86	3.40	2.35	0.69	1.57	5.76	0.43
267	1	1323	7	10-11h	1100	53.99	3.27	2.20	0.67	1.78	7.72	0.46
268	1	2254	7	12-14h	1151	53.99	3.13	2.09	0.67	1.60	5.98	0.43
278	1	648	7	17-17h	1572	46.35	2.29	1.48	0.65	1.97	7.72	0.14
279	1	774	7	17-18h	1385	46.35	2.60	1.66	0.64	2.08	10.70	0.17
286	1	601	7	20-21h	645	59.18	5.58	3.79	0.68	1.29	4.87	0.43
303	2	657	1	6-7h	1733	60.90	2.08	1.44	0.69	1.92	7.51	0.07
305	2	501	1	7-8h	1793	48.91	2.01	1.29	0.64	2.21	10.14	0.07
316	2	740	1	10-11h	1701	63.74	2.12	1.57	0.74	2.63	14.16	0.05
318	2	651	1	11-12h	1600	61.72	2.25	1.71	0.76	2.65	13.70	0.12
321	2	2151	1	12-14h	1585	62.70	2.27	1.64	0.72	2.23	10.95	0.09
323	2	1144	1	14-15h	1711	60.85	2.10	1.46	0.69	2.09	9.14	0.11
329	2	601	1	16-16h	1891	48.78	1.90	1.26	0.66	2.75	15.34	0.11
362	2	4401	2	10-12h	1604	62.47	2.24	1.65	0.74	2.39	12.00	0.11
363	2	2201	2	12-14h	1674	62.13	2.15	1.57	0.73	2.53	13.72	0.10
376	2	651	2	16-16h	2004	47.88	1.80	1.04	0.58	2.38	12.37	0.10
390	2	572	2	20-20h	1091	66.67	3.30	2.65	0.80	2.22	10.35	0.04
419	2	551	3	9-9h	1653	60.14	2.18	1.55	0.71	2.16	9.67	0.09
421	2	1451	3	10-11h	1612	62.25	2.23	1.64	0.73	2.40	11.76	0.09

L - LaneID; D-Day; mTH - mean Headway; sTH - st.dev. of Headway.

Continued...

ID	L	Size	D	Hour	Flow	Speed	mTH	sTH	CV	Sk.	Kur.	pLGV
431	2	551	3	18-18h	1808	48.04	1.99	1.30	0.65	2.07	8.46	0.07
461	2	643	4	7-8h	909	70.01	3.96	3.71	0.94	2.72	14.21	0.02
464	2	551	4	8-9h	1518	67.09	2.37	1.95	0.82	2.74	14.06	0.02
468	2	551	4	9-10h	1774	64.51	2.03	1.66	0.82	3.98	28.83	0.02
479	2	1801	4	11-12h	1933	55.84	1.86	1.28	0.69	2.91	17.30	0.03
480	2	851	4	12-13h	1777	56.95	2.03	1.52	0.75	3.17	19.56	0.02
483	2	2078	4	14-15h	1619	66.62	2.22	1.63	0.73	2.28	10.35	0.01
489	2	663	4	17-18h	1354	65.93	2.66	2.00	0.75	2.46	11.06	0.02
494	2	544	4	19-19h	918	67.24	3.92	3.73	0.95	2.38	9.96	0.01
514	2	586	5	9-9h	1068	66.97	3.37	2.91	0.86	3.07	17.53	0.02
517	2	885	5	10-11h	1328	62.82	2.71	2.03	0.75	2.53	11.68	0.02
519	2	1101	5	11-12h	1533	62.19	2.35	1.57	0.67	2.82	17.72	0.01
521	2	1151	5	12-13h	1530	62.44	2.35	1.76	0.75	2.98	18.98	0.02
522	2	1368	5	13-14h	1567	64.90	2.30	1.73	0.75	2.54	13.32	0.02
525	2	951	5	15-15h	1701	63.96	2.12	1.50	0.71	2.61	12.50	0.02
527	2	1351	5	15-16h	1786	64.62	2.02	1.40	0.69	2.73	14.59	0.01
533	2	2151	5	17-18h	1699	64.86	2.12	1.40	0.66	2.07	9.19	0.02
537	2	917	5	19-20h	1256	66.52	2.87	2.29	0.80	2.13	8.63	0.01
539	2	597	5	20-21h	1032	67.77	3.49	3.14	0.90	2.85	16.66	0.01
559	2	514	6	6-7h	1519	62.22	2.37	1.84	0.78	2.82	14.61	0.04
567	2	801	6	10-11h	1502	63.14	2.40	1.78	0.74	2.25	10.68	0.10
570	2	1851	6	11-13h	1590	62.65	2.26	1.62	0.71	1.97	7.81	0.09
577	2	501	6	14-14h	1604	62.64	2.24	1.51	0.67	2.00	8.55	0.10
578	2	899	6	14-15h	1622	62.47	2.22	1.66	0.75	2.86	16.19	0.09
582	2	851	6	15-16h	1685	61.20	2.14	1.52	0.71	2.24	9.70	0.08
588	2	763	6	19-20h	908	66.71	3.96	3.93	0.99	3.15	17.74	0.06
624	2	981	7	10-11h	1338	61.88	2.69	2.01	0.75	2.21	10.80	0.08
626	2	1417	7	11-12h	1392	61.05	2.59	1.83	0.71	2.88	19.25	0.12
627	2	1001	7	12-13h	1472	61.05	2.44	1.71	0.70	2.27	10.37	0.11
632	2	785	7	13-14h	1567	61.24	2.30	1.54	0.67	2.14	10.37	0.13
634	2	501	7	14-14h	1565	62.66	2.30	1.67	0.73	2.11	8.95	0.10
636	2	700	7	15-15h	1714	60.52	2.10	1.46	0.69	1.90	7.22	0.09
639	2	551	7	15-16h	1866	56.08	1.93	1.25	0.65	1.95	7.70	0.07
649	2	511	7	19-20h	1025	65.79	3.51	2.78	0.79	1.82	6.69	0.04
686	3	1254	1	13-14h	1308	70.23	2.75	3.13	1.14	2.95	14.38	0.00
692	3	601	1	14-15h	1710	66.69	2.11	2.75	1.31	4.38	29.00	0.00
698	3	901	1	16-17h	1787	52.39	2.02	2.10	1.04	4.70	40.35	0.00
733	3	551	2	10-10h	1274	70.17	2.83	3.53	1.25	3.16	15.15	0.00
742	3	1547	2	13-14h	1513	69.87	2.38	2.62	1.10	3.03	14.87	0.00
757	3	601	2	17-17h	1438	52.74	2.50	2.70	1.08	2.69	10.91	0.00
782	3	793	3	8-8h	1463	52.91	2.46	2.96	1.20	3.55	19.44	0.00

L - LaneID; D-Day; mTH - mean Headway; sTH - st.dev. of Headway.

Continued...

ID	L	Size	D	Hour	Flow	Speed	mTH	sTH	CV	Sk.	Kur.	pLGV
787	3	1328	3	9-10h	1339	70.21	2.69	3.46	1.29	4.91	43.69	0.00
800	3	651	3	14-14h	2004	49.06	1.80	1.78	0.99	4.89	34.41	0.00
803	3	501	3	14-15h	1911	50.42	1.88	1.79	0.95	4.04	25.56	0.00
816	3	1151	3	17-17h	1870	51.03	1.93	2.00	1.04	4.55	33.29	0.00
818	3	554	3	18-19h	1328	52.46	2.71	3.10	1.14	2.76	11.17	0.00
823	3	751	3	19-20h	1336	67.19	2.69	3.15	1.17	3.21	15.63	0.00
825	3	786	3	20-21h	1088	64.93	3.31	4.05	1.22	3.09	14.74	0.00
848	3	851	4	10-10h	1491	68.51	2.41	2.98	1.23	3.91	22.89	0.00
855	3	501	4	12-13h	1281	61.90	2.81	3.48	1.24	3.15	15.60	0.00
860	3	1551	4	14-16h	906	75.48	3.97	4.97	1.25	2.99	14.59	0.00
876	3	701	5	10-11h	956	69.94	3.77	4.20	1.12	2.85	13.72	0.00
878	3	716	5	11-12h	1244	68.30	2.89	3.82	1.32	5.05	37.88	0.00
892	3	1030	5	15-16h	1433	71.80	2.51	2.84	1.13	3.01	14.00	0.00
896	3	1650	5	17-18h	1446	71.43	2.49	2.90	1.16	3.80	24.96	0.00
900	3	794	5	18-19h	1042	72.09	3.45	4.39	1.27	3.59	22.45	0.00
901	3	700	5	19-20h	789	72.09	4.57	6.04	1.32	2.90	13.35	0.00
922	3	604	6	9-9h	1035	54.39	3.48	4.54	1.30	3.10	15.47	0.00
929	3	801	6	12-12h	1225	70.14	2.94	3.53	1.20	3.45	19.71	0.00
931	3	1401	6	13-14h	1258	70.33	2.86	3.42	1.19	3.59	22.27	0.00
939	3	611	6	16-16h	1527	60.24	2.36	2.60	1.10	3.48	20.21	0.00
941	3	601	6	16-17h	1333	54.26	2.70	2.97	1.10	2.79	12.42	0.00
962	3	588	7	6-7h	1414	67.46	2.55	2.75	1.08	3.12	14.57	0.00
977	3	1051	7	10-11h	1039	69.52	3.46	4.14	1.20	2.94	13.77	0.00
979	3	1063	7	11-12h	1161	68.57	3.10	3.61	1.16	3.38	18.99	0.00
981	3	901	7	12-13h	1108	67.93	3.25	3.80	1.17	2.93	13.66	0.00
997	3	601	7	17-17h	1614	51.41	2.23	2.24	1.00	2.85	12.51	0.00

L - LaneID; D-Day; mTH - mean Headway; sTH - st.dev. of Headway.

Appendix B

Selected Test Results

Table B.1: Results of gamma-SPM.

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
63	1	1725	0.00	0.88	9.48	0.11	1.08	0.03	0.84	0.33	0.91	1.02	2.72
69	1	1158	0.11	0.46	11.17	0.11	1.16	0.02	0.96	0.27	0.96	1.25	1.55
70	1	1223	0.17	0.50	8.46	0.16	1.92	0.01	0.92	0.24	0.98	1.36	1.45
71	1	1271	0.25	0.49	10.59	0.13	1.00	0.02	0.97	0.20	0.99	1.36	1.58
76	1	1483	0.39	0.56	8.68	0.16	0.68	0.02	0.97	0.23	0.98	1.40	1.65
79	1	1444	0.42	0.50	9.27	0.15	1.00	0.02	1.00	0.14	1.00	1.40	1.52
80	1	1192	0.28	0.40	15.91	0.08	1.08	0.03	0.57	0.36	0.88	1.23	1.59
81	1	873	0.00	0.33	7.13	0.17	1.86	0.04	0.16	1.35	0.22	1.19	0.89
86	1	662	0.00	0.27	3.78	0.48	0.86	0.04	0.31	1.22	0.26	1.83	0.52
98	1	1195	0.16	0.47	9.89	0.13	1.92	0.01	0.82	0.34	0.91	1.28	1.47
99	1	1222	0.17	0.48	8.90	0.14	1.45	0.01	0.98	0.31	0.93	1.29	1.44
103	1	1477	0.25	0.61	10.46	0.12	1.61	0.01	0.99	0.22	0.99	1.28	1.97
112	1	901	0.00	0.35	4.97	0.26	0.64	0.03	0.63	0.63	0.62	1.27	0.79
120	1	573	0.00	0.22	3.42	0.57	0.98	0.04	0.21	1.38	0.21	1.95	0.41
128	1	1216	0.26	0.46	8.64	0.17	1.40	0.02	0.81	0.22	0.99	1.43	1.35
139	1	1559	0.56	0.52	7.85	0.19	0.72	0.03	0.73	0.36	0.89	1.53	1.45
151	1	1584	0.38	0.63	8.52	0.16	0.98	0.02	0.85	0.38	0.87	1.36	1.82
162	1	449	0.00	0.16	4.54	0.42	0.95	0.03	0.59	0.66	0.59	1.89	0.34
164	1	702	0.00	0.29	2.31	0.89	1.11	0.03	0.60	0.71	0.55	2.06	0.44
167	1	1012	0.00	0.39	4.29	0.24	0.76	0.04	0.13	1.60	0.16	1.01	0.80
168	1	1168	0.06	0.45	13.74	0.07	1.33	0.03	0.70	0.53	0.71	1.01	1.67
169	1	1203	0.13	0.48	10.56	0.12	1.72	0.02	0.57	0.49	0.76	1.23	1.56
170	1	1515	0.42	0.53	9.90	0.13	1.15	0.02	0.90	0.26	0.96	1.33	1.66
173	1	1163	0.24	0.42	10.86	0.12	0.90	0.03	0.83	0.38	0.87	1.33	1.38

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway

Continued...

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
175	1	1182	0.00	0.49	4.93	0.23	0.68	0.04	0.16	1.74	0.13	1.12	1.10
177	1	1039	0.00	0.40	6.34	0.17	0.92	0.02	0.97	0.33	0.91	1.07	1.02
178	1	943	0.00	0.36	15.64	0.07	2.44	0.02	0.63	0.65	0.60	1.04	1.41
181	1	886	0.00	0.37	5.90	0.25	1.30	0.02	0.92	0.32	0.92	1.47	0.89
186	1	552	0.01	0.19	8.10	0.15	2.19	0.03	0.36	0.59	0.65	1.24	0.53
187	1	457	0.00	0.15	9.35	0.13	1.27	0.02	0.84	0.18	0.99	1.22	0.46
189	1	196	0.00	0.06	25.07	0.05	2.93	0.03	0.62	0.34	0.91	1.26	0.29
200	1	999	0.22	0.38	8.13	0.20	0.81	0.02	0.96	0.28	0.96	1.64	1.08
203	1	1149	0.00	0.48	8.02	0.14	2.10	0.02	0.27	0.92	0.41	1.12	1.36
204	1	1102	0.04	0.43	11.62	0.09	3.19	0.02	0.71	0.44	0.81	1.06	1.46
205	1	1142	0.00	0.46	7.55	0.14	0.78	0.03	0.84	0.66	0.59	1.08	1.26
206	1	1147	0.00	0.44	21.77	0.04	3.30	0.01	0.88	0.26	0.96	0.86	2.03
208	1	1107	0.14	0.43	11.09	0.12	1.11	0.01	1.00	0.14	1.00	1.33	1.44
209	1	1027	0.23	0.38	6.93	0.23	0.80	0.02	0.96	0.30	0.94	1.61	1.01
210	1	912	0.00	0.34	11.07	0.10	2.31	0.04	0.04	1.73	0.13	1.07	1.14
211	1	777	0.00	0.31	3.87	0.41	0.81	0.04	0.42	0.62	0.63	1.58	0.61
234	1	1561	0.36	0.60	9.57	0.14	1.29	0.01	0.99	0.17	1.00	1.30	1.86
236	1	1168	0.00	0.48	5.87	0.19	0.97	0.03	0.22	0.92	0.40	1.10	1.17
238	1	1169	0.00	0.46	11.96	0.08	10.35	0.01	0.70	0.47	0.78	0.96	1.60
239	1	1277	0.17	0.54	8.87	0.15	0.79	0.02	0.96	0.31	0.93	1.36	1.60
241	1	1315	0.26	0.55	8.87	0.17	0.81	0.03	0.85	0.45	0.79	1.50	1.65
257	1	1535	0.41	0.65	8.04	0.19	0.90	0.02	0.93	0.19	0.99	1.54	1.85
266	1	1058	0.13	0.41	11.54	0.11	1.76	0.02	0.93	0.26	0.97	1.31	1.38
267	1	1100	0.24	0.44	7.68	0.21	1.14	0.02	0.81	0.45	0.80	1.65	1.21
268	1	1151	0.09	0.47	10.18	0.12	2.02	0.01	0.92	0.26	0.97	1.24	1.49
278	1	1572	0.13	0.69	17.79	0.06	1.21	0.03	0.66	0.36	0.89	1.06	2.91
279	1	1385	0.24	0.60	9.23	0.15	0.97	0.01	1.00	0.13	1.00	1.41	1.81
286	1	645	0.00	0.23	8.59	0.15	1.77	0.07	0.00	2.39	0.06	1.31	0.68
303	2	1733	0.00	0.72	9.69	0.08	1.08	0.05	0.11	1.44	0.19	0.76	2.25
305	2	1793	0.18	0.81	13.63	0.08	1.20	0.02	0.91	0.28	0.95	1.04	2.97
316	2	1701	0.30	0.65	10.12	0.11	0.98	0.02	0.91	0.38	0.87	1.09	2.06
318	2	1600	0.30	0.59	9.45	0.12	0.93	0.02	0.93	0.27	0.96	1.11	1.81
321	2	1585	0.23	0.61	10.82	0.10	2.19	0.01	0.97	0.21	0.99	1.05	2.01
323	2	1711	0.23	0.69	11.28	0.09	1.38	0.02	0.88	0.29	0.95	1.03	2.30
329	2	1891	0.59	0.68	6.95	0.20	0.76	0.03	0.80	0.34	0.90	1.37	1.79
362	2	1604	0.28	0.60	10.68	0.10	3.74	0.01	0.57	0.52	0.73	1.09	1.98
363	2	1674	0.15	0.67	13.32	0.07	2.82	0.02	0.24	0.61	0.64	0.92	2.45
376	2	2004	0.38	0.93	11.26	0.11	0.97	0.02	0.81	0.32	0.92	1.19	3.11
390	2	1091	0.01	0.39	11.65	0.06	1.57	0.02	0.97	0.21	0.99	0.75	1.32
419	2	1653	0.34	0.62	8.68	0.13	0.82	0.02	1.00	0.14	1.00	1.17	1.82

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway
Continued...

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
421	2	1612	0.27	0.61	12.15	0.09	1.95	0.02	0.53	0.36	0.89	1.08	2.13
431	2	1808	0.43	0.71	9.82	0.13	0.84	0.01	1.00	0.09	1.00	1.24	2.21
461	2	909	0.16	0.30	8.30	0.14	1.18	0.04	0.20	1.01	0.35	1.17	0.85
464	2	1518	0.38	0.50	7.80	0.15	0.85	0.02	0.99	0.18	0.99	1.19	1.40
468	2	1774	0.68	0.49	6.30	0.23	0.62	0.02	0.98	0.36	0.89	1.42	1.24
479	2	1933	0.38	0.78	10.14	0.11	2.50	0.02	0.68	0.68	0.57	1.13	2.50
480	2	1777	0.60	0.57	7.32	0.19	0.99	0.03	0.33	0.92	0.40	1.37	1.53
483	2	1619	0.25	0.62	10.57	0.10	2.40	0.02	0.63	0.40	0.84	1.05	2.02
489	2	1354	0.50	0.45	6.11	0.27	0.77	0.03	0.48	0.81	0.47	1.65	1.12
494	2	918	0.46	0.24	5.12	0.35	0.82	0.02	0.88	0.33	0.91	1.77	0.55
514	2	1068	0.37	0.35	7.28	0.23	0.84	0.04	0.43	0.47	0.78	1.65	0.95
517	2	1328	0.27	0.51	9.90	0.14	1.45	0.03	0.52	0.78	0.50	1.35	1.61
519	2	1533	0.16	0.68	11.28	0.11	1.61	0.02	0.70	0.35	0.90	1.19	2.29
521	2	1530	0.49	0.53	7.23	0.20	1.26	0.01	0.97	0.30	0.94	1.44	1.41
522	2	1567	0.29	0.58	10.54	0.11	1.70	0.02	0.88	0.31	0.93	1.13	1.89
525	2	1701	0.48	0.62	7.68	0.18	1.06	0.02	0.95	0.35	0.90	1.35	1.72
527	2	1786	0.31	0.73	11.38	0.10	1.50	0.02	0.67	0.41	0.84	1.12	2.47
533	2	1699	0.24	0.71	11.36	0.10	2.37	0.01	0.99	0.13	1.00	1.10	2.39
537	2	1256	0.20	0.45	11.83	0.10	1.40	0.03	0.50	0.80	0.48	1.13	1.55
539	2	1032	0.18	0.34	13.11	0.08	1.21	0.03	0.80	0.25	0.97	1.09	1.23
559	2	1519	0.20	0.59	12.80	0.08	1.14	0.05	0.17	0.78	0.50	1.06	2.13
567	2	1502	0.21	0.56	10.25	0.10	1.34	0.01	1.00	0.17	1.00	1.03	1.81
570	2	1590	0.15	0.63	12.33	0.08	2.50	0.02	0.56	0.62	0.63	0.95	2.22
577	2	1604	0.18	0.67	10.69	0.10	1.00	0.02	0.98	0.23	0.98	1.07	2.18
578	2	1622	0.00	0.66	11.47	0.06	1.68	0.02	0.78	0.58	0.67	0.74	2.25
582	2	1685	0.28	0.66	11.92	0.09	1.44	0.02	0.92	0.28	0.96	1.09	2.27
588	2	908	0.18	0.29	11.77	0.10	1.30	0.04	0.28	1.19	0.27	1.15	0.99
624	2	1338	0.16	0.51	12.80	0.08	1.61	0.03	0.56	0.46	0.79	1.08	1.82
626	2	1392	0.24	0.57	7.56	0.18	1.25	0.01	0.99	0.26	0.96	1.33	1.56
627	2	1472	0.23	0.60	8.93	0.14	1.39	0.01	0.99	0.24	0.97	1.22	1.79
632	2	1567	0.17	0.65	10.31	0.11	1.30	0.02	0.92	0.17	1.00	1.08	2.10
634	2	1565	0.29	0.59	9.33	0.12	0.91	0.03	0.89	0.36	0.89	1.15	1.80
636	2	1714	0.35	0.64	9.03	0.13	0.97	0.02	0.83	0.22	0.98	1.14	1.93
639	2	1866	0.00	0.83	20.58	0.04	1.62	0.03	0.60	0.61	0.64	0.75	3.77
649	2	1025	0.07	0.37	8.20	0.12	1.18	0.04	0.36	0.61	0.63	1.02	1.05
686	3	1308	0.47	0.30	10.04	0.10	1.44	0.03	0.12	0.98	0.37	1.03	0.96
692	3	1710	0.72	0.27	8.02	0.14	1.10	0.03	0.64	0.78	0.49	1.10	0.78
698	3	1787	0.64	0.41	10.83	0.11	1.50	0.02	0.65	0.41	0.84	1.14	1.34
733	3	1274	0.52	0.27	10.48	0.10	1.03	0.03	0.63	0.63	0.62	1.04	0.86
742	3	1513	0.53	0.35	10.21	0.10	2.20	0.02	0.35	0.87	0.43	1.03	1.11

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway

Continued...

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
757	3	1438	0.58	0.31	10.77	0.11	0.95	0.02	0.86	0.35	0.89	1.16	1.01
782	3	1463	0.57	0.31	12.51	0.09	1.14	0.04	0.18	1.26	0.24	1.09	1.11
787	3	1339	0.54	0.29	8.04	0.14	1.49	0.03	0.12	1.89	0.11	1.12	0.82
800	3	2004	0.75	0.43	8.98	0.14	0.87	0.05	0.12	1.55	0.17	1.23	1.27
803	3	1911	0.77	0.38	8.14	0.16	0.82	0.05	0.21	1.44	0.19	1.30	1.07
816	3	1870	0.67	0.43	11.19	0.10	1.45	0.04	0.09	2.15	0.08	1.16	1.43
818	3	1328	0.58	0.27	11.38	0.10	0.95	0.04	0.27	0.86	0.44	1.17	0.92
823	3	1336	0.60	0.27	6.77	0.18	1.01	0.04	0.28	1.05	0.33	1.25	0.70
825	3	1088	0.59	0.21	5.59	0.25	0.86	0.03	0.63	0.79	0.49	1.39	0.50
848	3	1491	0.67	0.28	6.92	0.18	1.05	0.03	0.48	1.46	0.19	1.24	0.73
855	3	1281	0.56	0.25	10.03	0.11	0.83	0.03	0.88	0.39	0.86	1.08	0.80
860	3	906	0.43	0.20	6.67	0.17	1.47	0.03	0.08	2.02	0.09	1.14	0.52
876	3	956	0.46	0.23	8.67	0.16	1.07	0.03	0.52	0.67	0.59	1.40	0.66
878	3	1244	0.61	0.25	7.30	0.19	1.20	0.04	0.33	1.14	0.29	1.38	0.69
892	3	1433	0.55	0.31	9.89	0.11	1.30	0.02	0.82	0.53	0.72	1.07	0.98
896	3	1446	0.59	0.30	8.41	0.14	1.69	0.02	0.40	1.02	0.34	1.14	0.88
900	3	1042	0.49	0.22	7.14	0.16	1.15	0.03	0.59	0.87	0.43	1.16	0.59
901	3	789	0.43	0.17	7.41	0.16	1.20	0.05	0.07	1.33	0.22	1.16	0.46
922	3	1035	0.47	0.22	13.21	0.08	1.29	0.05	0.08	1.08	0.32	1.04	0.79
929	3	1225	0.51	0.27	8.33	0.14	0.96	0.03	0.38	0.82	0.47	1.14	0.78
931	3	1258	0.51	0.28	8.69	0.13	1.61	0.02	0.54	0.84	0.45	1.11	0.81
939	3	1527	0.55	0.35	10.41	0.11	1.06	0.03	0.48	0.55	0.70	1.10	1.14
941	3	1333	0.51	0.30	10.70	0.10	1.03	0.03	0.80	0.33	0.91	1.10	0.99
962	3	1414	0.58	0.32	8.98	0.14	1.01	0.02	0.85	0.49	0.76	1.24	0.95
977	3	1039	0.56	0.21	7.05	0.20	1.18	0.02	0.61	0.67	0.59	1.38	0.56
979	3	1161	0.54	0.25	7.58	0.17	1.19	0.03	0.36	0.61	0.64	1.33	0.70
981	3	1108	0.58	0.22	7.02	0.20	0.98	0.03	0.32	0.90	0.42	1.39	0.59
997	3	1614	0.60	0.37	12.83	0.09	1.01	0.03	0.51	0.53	0.71	1.15	1.32

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway

Table B.2: Results of gamma-GQM.

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
63	1	1724.62	0	0.85	12.38	0.07	0.57	0.03	0.9	0.2	0.99	0.91	2.98
69	1	1157.53	0.15	0.46	11.27	0.11	0.95	0.02	0.96	0.27	0.96	1.27	1.56
70	1	1222.53	0.22	0.5	8.57	0.16	1.78	0.01	0.92	0.24	0.98	1.39	1.46
71	1	1270.64	0.29	0.49	10.64	0.13	0.61	0.02	0.97	0.2	0.99	1.37	1.58
76	1	1483.39	0.43	0.56	8.73	0.16	0.73	0.02	0.97	0.23	0.98	1.41	1.65

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway
Continued...

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
79	1	1444.1	0.45	0.5	9.3	0.15	0.59	0.02	1	0.14	1	1.4	1.53
80	1	1191.71	0.29	0.4	15.92	0.08	0.81	0.03	0.57	0.36	0.88	1.24	1.59
81	1	873.1	0	0.35	5.61	0.21	0.86	0.05	0.05	2.04	0.09	1.2	0.83
86	1	661.58	0	0.26	3.86	0.42	0.58	0.04	0.48	0.98	0.37	1.6	0.51
98	1	1194.67	0.2	0.47	9.96	0.13	1.51	0.01	0.82	0.34	0.91	1.3	1.48
99	1	1222.27	0.21	0.48	8.99	0.15	1.08	0.01	0.98	0.31	0.93	1.31	1.45
103	1	1477.28	0.3	0.61	10.53	0.12	1.04	0.01	0.99	0.22	0.99	1.29	1.97
112	1	900.58	0.03	0.34	12.43	0.09	0.75	0.03	0.8	0.33	0.91	1.16	1.2
120	1	573.24	0	0.22	4.19	0.39	0.72	0.05	0.07	1.68	0.14	1.64	0.44
128	1	1216.22	0.3	0.46	8.69	0.17	1.31	0.02	0.81	0.22	0.99	1.45	1.36
139	1	1558.82	0.6	0.52	7.87	0.2	0.52	0.03	0.73	0.36	0.89	1.54	1.45
151	1	1584.33	0.44	0.62	8.56	0.16	0.67	0.02	0.85	0.38	0.87	1.37	1.83
162	1	448.99	0.03	0.16	8.28	0.22	0.77	0.04	0.43	1.03	0.34	1.84	0.45
164	1	702.05	0.1	0.27	4.6	0.39	0.91	0.04	0.26	1.17	0.28	1.79	0.58
167	1	1012.32	0	0.36	20.33	0.04	1.31	0.04	0.14	1.32	0.23	0.81	1.64
168	1	1168.34	0.08	0.45	13.86	0.07	0.9	0.03	0.7	0.53	0.71	1.03	1.68
169	1	1202.88	0.16	0.48	10.67	0.12	0.88	0.02	0.57	0.49	0.76	1.25	1.57
170	1	1515.34	0.45	0.53	9.93	0.13	0.68	0.02	0.9	0.26	0.96	1.34	1.66
173	1	1163.36	0.26	0.42	10.9	0.12	0.74	0.03	0.83	0.38	0.87	1.34	1.38
175	1	1182.06	0.14	0.46	12.6	0.09	1.07	0.03	0.54	0.62	0.63	1.2	1.64
177	1	1039.2	0	0.41	6.89	0.15	0.76	0.03	0.6	0.63	0.62	1.06	1.09
178	1	943.22	0	0.37	8.15	0.13	0.83	0.02	0.63	0.58	0.66	1.09	1.05
181	1	885.74	0	0.37	5.91	0.23	0.99	0.02	0.92	0.33	0.92	1.34	0.89
186	1	552.37	0	0.19	7.81	0.15	1.6	0.03	0.36	0.59	0.66	1.18	0.52
187	1	456.68	0.02	0.15	10.95	0.12	1.09	0.02	0.86	0.19	0.99	1.3	0.49
189	1	195.86	0	0.06	25.16	0.05	1.56	0.03	0.62	0.34	0.91	1.27	0.29
200	1	998.68	0.27	0.38	8.2	0.2	0.62	0.02	0.95	0.28	0.96	1.67	1.08
203	1	1148.95	0.12	0.47	12.13	0.1	1.48	0.01	0.96	0.27	0.96	1.25	1.63
204	1	1101.98	0	0.43	11.38	0.08	2.29	0.02	0.72	0.44	0.81	0.94	1.45
205	1	1141.99	0	0.45	28.06	0.03	1.21	0.04	0.39	0.5	0.74	0.92	2.37
206	1	1147.15	0	0.45	15.93	0.05	3.54	0.02	0.43	0.9	0.41	0.86	1.8
208	1	1106.66	0.18	0.43	11.16	0.12	0.92	0.01	1	0.14	1	1.35	1.45
209	1	1027.33	0.28	0.38	6.99	0.23	0.73	0.02	0.96	0.3	0.94	1.64	1.02
210	1	912.38	0	0.35	5.9	0.19	0.85	0.03	0.11	1.39	0.21	1.1	0.85
211	1	777.31	0.13	0.31	4.45	0.42	0.66	0.03	0.57	0.5	0.75	1.86	0.66
234	1	1561.17	0.4	0.6	9.62	0.14	0.88	0.01	0.99	0.17	1	1.31	1.87
236	1	1168.04	0.21	0.46	8.81	0.15	0.92	0.01	0.98	0.19	0.99	1.36	1.37
238	1	1168.96	0.04	0.46	12.95	0.08	3.42	0.01	0.7	0.47	0.78	1	1.67
239	1	1277.09	0.24	0.54	8.96	0.16	0.72	0.02	0.96	0.31	0.93	1.39	1.6
241	1	1314.53	0.32	0.55	8.98	0.17	0.66	0.03	0.85	0.45	0.79	1.52	1.66

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway

Continued...

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
257	1	1534.82	0.49	0.65	8.12	0.19	0.59	0.02	0.93	0.19	0.99	1.56	1.86
266	1	1058.18	0.15	0.41	11.61	0.11	0.71	0.02	0.93	0.26	0.97	1.33	1.39
267	1	1100.12	0.3	0.44	7.78	0.22	1.17	0.02	0.81	0.45	0.79	1.68	1.22
268	1	1150.85	0.13	0.47	10.32	0.12	1.64	0.01	0.92	0.26	0.97	1.27	1.5
278	1	1572.01	0.16	0.69	17.89	0.06	0.94	0.03	0.66	0.36	0.89	1.07	2.91
279	1	1385.06	0.3	0.6	9.34	0.15	0.85	0.01	1	0.13	1	1.43	1.82
286	1	644.87	0	0.25	3.57	0.42	0.63	0.06	0.02	1.7	0.13	1.5	0.46
303	2	1733.07	0.24	0.68	14.21	0.07	0.96	0.02	0.99	0.16	1	0.96	2.57
305	2	1793.08	0.23	0.8	13.67	0.08	0.86	0.02	0.91	0.28	0.95	1.05	2.98
316	2	1700.84	0.34	0.65	10.16	0.11	0.85	0.02	0.91	0.38	0.87	1.1	2.06
318	2	1600.43	0.33	0.59	9.49	0.12	0.68	0.02	0.93	0.27	0.96	1.12	1.81
321	2	1585.27	0.26	0.61	10.87	0.1	1.74	0.01	0.97	0.21	0.99	1.06	2.01
323	2	1710.95	0.27	0.69	11.34	0.09	1.04	0.02	0.88	0.29	0.95	1.04	2.31
329	2	1890.75	0.64	0.68	6.99	0.2	0.6	0.03	0.8	0.34	0.9	1.38	1.8
362	2	1604.03	0.31	0.6	10.72	0.1	2.87	0.01	0.57	0.51	0.73	1.1	1.98
363	2	1674.34	0.18	0.67	13.33	0.07	1.83	0.02	0.24	0.6	0.65	0.93	2.45
376	2	2003.76	0.45	0.93	11.35	0.11	0.87	0.02	0.81	0.32	0.92	1.2	3.13
390	2	1091.13	0	0.39	6.24	0.12	0.73	0.03	0.84	0.3	0.94	0.75	0.97
419	2	1653.38	0.38	0.62	8.73	0.13	0.63	0.02	1	0.14	1	1.18	1.82
421	2	1612.01	0.3	0.61	12.19	0.09	1.56	0.02	0.53	0.36	0.89	1.08	2.14
431	2	1807.75	0.47	0.71	9.87	0.13	0.78	0.01	1	0.09	1	1.24	2.22
461	2	908.71	0.18	0.3	8.3	0.14	1.03	0.04	0.2	1.01	0.35	1.19	0.85
464	2	1518.2	0.41	0.5	7.83	0.15	0.58	0.02	0.99	0.18	0.99	1.19	1.41
468	2	1773.86	0.7	0.49	6.32	0.23	0.55	0.02	0.98	0.36	0.89	1.42	1.24
479	2	1933.13	0.43	0.78	10.18	0.11	1.48	0.02	0.68	0.68	0.58	1.14	2.5
480	2	1776.74	0.63	0.57	7.34	0.19	0.67	0.03	0.33	0.92	0.4	1.37	1.54
483	2	1618.93	0.28	0.62	10.61	0.1	1.51	0.02	0.64	0.4	0.85	1.06	2.02
489	2	1354.42	0.55	0.45	6.14	0.27	0.6	0.03	0.48	0.8	0.48	1.66	1.12
494	2	917.96	0.48	0.24	5.13	0.35	0.55	0.02	0.88	0.33	0.91	1.78	0.55
514	2	1068.04	0.4	0.35	7.3	0.23	0.69	0.04	0.43	0.47	0.78	1.66	0.95
517	2	1327.72	0.31	0.51	9.95	0.14	0.93	0.03	0.52	0.78	0.5	1.36	1.62
519	2	1533.38	0.22	0.68	11.34	0.11	1.28	0.02	0.7	0.34	0.9	1.21	2.29
521	2	1530.2	0.53	0.53	7.26	0.2	0.65	0.01	0.97	0.3	0.94	1.45	1.42
522	2	1567.39	0.33	0.58	10.58	0.11	1.12	0.02	0.88	0.31	0.93	1.14	1.89
525	2	1701.03	0.53	0.62	7.72	0.18	0.93	0.02	0.95	0.35	0.9	1.36	1.73
527	2	1786.31	0.35	0.73	11.44	0.1	1.42	0.02	0.67	0.41	0.84	1.13	2.48
533	2	1699.07	0.28	0.71	11.43	0.1	1.85	0.01	0.99	0.13	1	1.11	2.4
537	2	1255.55	0.22	0.45	11.85	0.1	0.97	0.03	0.5	0.8	0.48	1.14	1.55
539	2	1032.36	0.19	0.34	13.12	0.08	0.91	0.03	0.8	0.25	0.97	1.1	1.23
559	2	1519.27	0.23	0.59	12.85	0.08	0.76	0.05	0.17	0.78	0.5	1.07	2.13

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway

Continued...

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
567	2	1501.52	0.24	0.56	10.3	0.1	0.83	0.01	1	0.17	1	1.04	1.81
570	2	1590.35	0.18	0.63	12.37	0.08	1.9	0.02	0.56	0.61	0.63	0.96	2.22
577	2	1603.92	0.22	0.67	10.79	0.1	0.73	0.02	0.98	0.23	0.98	1.08	2.19
578	2	1621.62	0.05	0.66	13.19	0.06	1.14	0.02	0.7	0.64	0.61	0.78	2.39
582	2	1685	0.32	0.66	11.96	0.09	0.9	0.02	0.92	0.28	0.96	1.1	2.27
588	2	908.36	0.19	0.29	11.78	0.1	0.86	0.04	0.28	1.19	0.27	1.16	0.99
624	2	1338.19	0.18	0.51	12.84	0.08	1.15	0.03	0.56	0.46	0.79	1.09	1.83
626	2	1392.02	0.3	0.57	7.66	0.18	0.92	0.01	0.99	0.26	0.96	1.36	1.57
627	2	1472.45	0.28	0.6	9	0.14	1	0.01	0.99	0.24	0.97	1.24	1.79
632	2	1566.92	0.22	0.65	10.41	0.11	0.8	0.02	0.92	0.17	1	1.1	2.11
634	2	1564.55	0.33	0.59	9.37	0.12	0.77	0.03	0.89	0.36	0.89	1.16	1.8
636	2	1714.46	0.39	0.64	9.07	0.13	0.78	0.02	0.83	0.22	0.98	1.15	1.93
639	2	1866.08	0	0.83	20.69	0.03	1.22	0.03	0.6	0.61	0.64	0.72	3.77
649	2	1025.43	0.1	0.37	8.26	0.13	0.9	0.04	0.36	0.61	0.63	1.05	1.06
686	3	1308.1	0.48	0.3	10.04	0.1	1.27	0.03	0.12	0.97	0.37	1.03	0.96
692	3	1710.08	0.73	0.27	8.02	0.14	0.77	0.03	0.64	0.78	0.49	1.1	0.78
698	3	1786.53	0.64	0.41	10.83	0.11	0.77	0.02	0.65	0.41	0.84	1.14	1.34
733	3	1274.15	0.53	0.27	10.48	0.1	0.76	0.03	0.63	0.63	0.62	1.04	0.86
742	3	1512.91	0.53	0.35	10.21	0.1	1.19	0.02	0.35	0.87	0.43	1.03	1.11
757	3	1437.96	0.59	0.31	10.78	0.11	0.77	0.02	0.86	0.35	0.89	1.16	1.01
782	3	1462.5	0.57	0.31	12.52	0.09	0.93	0.04	0.18	1.26	0.24	1.09	1.11
787	3	1338.58	0.55	0.29	8.04	0.14	1.14	0.03	0.12	1.89	0.11	1.12	0.82
800	3	2004.09	0.76	0.43	8.98	0.14	0.74	0.05	0.12	1.55	0.17	1.23	1.28
803	3	1910.9	0.78	0.38	8.15	0.16	0.58	0.05	0.21	1.44	0.19	1.3	1.07
816	3	1869.64	0.68	0.43	11.2	0.1	1.09	0.04	0.09	2.15	0.08	1.16	1.43
818	3	1328.32	0.58	0.27	11.38	0.1	0.71	0.04	0.27	0.86	0.44	1.17	0.92
823	3	1336.25	0.61	0.27	6.77	0.19	0.77	0.04	0.28	1.05	0.33	1.25	0.7
825	3	1087.52	0.59	0.21	5.59	0.25	0.77	0.03	0.63	0.79	0.49	1.39	0.5
848	3	1491.08	0.67	0.28	6.92	0.18	0.67	0.03	0.48	1.46	0.19	1.24	0.73
855	3	1280.65	0.56	0.25	10.03	0.11	0.6	0.03	0.88	0.39	0.86	1.08	0.8
860	3	905.86	0.43	0.2	6.67	0.17	1.17	0.03	0.08	2.02	0.09	1.14	0.52
876	3	955.87	0.47	0.23	8.67	0.16	0.67	0.03	0.52	0.67	0.59	1.41	0.66
878	3	1243.73	0.62	0.25	7.3	0.19	0.64	0.04	0.33	1.14	0.29	1.39	0.69
892	3	1433.36	0.55	0.31	9.89	0.11	0.78	0.02	0.82	0.53	0.72	1.07	0.98
896	3	1445.7	0.59	0.3	8.41	0.14	1.12	0.02	0.4	1.02	0.35	1.14	0.88
900	3	1042.32	0.49	0.22	7.14	0.16	0.71	0.03	0.59	0.87	0.43	1.17	0.59
901	3	788.6	0.43	0.17	7.41	0.16	0.99	0.05	0.07	1.33	0.22	1.16	0.46
922	3	1035.31	0.47	0.22	13.21	0.08	0.84	0.05	0.08	1.08	0.32	1.04	0.79
929	3	1224.64	0.52	0.27	8.33	0.14	0.88	0.03	0.38	0.82	0.47	1.14	0.78
931	3	1257.71	0.52	0.28	8.7	0.13	1.07	0.02	0.54	0.84	0.45	1.11	0.81

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway

Continued...

ID	L	Fl	ϕ	λ	α	β	Ft	KSs	KSp	Ads	Adp	mFH	sFH
939	3	1527.09	0.56	0.35	10.41	0.11	0.68	0.03	0.48	0.55	0.7	1.1	1.14
941	3	1333.12	0.52	0.3	10.7	0.1	0.77	0.03	0.8	0.33	0.91	1.1	0.99
962	3	1413.96	0.59	0.32	8.98	0.14	0.72	0.02	0.85	0.49	0.76	1.24	0.95
977	3	1039.3	0.57	0.21	7.05	0.2	0.95	0.02	0.61	0.67	0.59	1.39	0.56
979	3	1161.19	0.55	0.25	7.58	0.18	0.88	0.03	0.36	0.61	0.64	1.33	0.7
981	3	1107.8	0.59	0.22	7.02	0.2	0.78	0.03	0.32	0.89	0.42	1.39	0.59
997	3	1614.15	0.6	0.37	12.83	0.09	0.92	0.03	0.51	0.53	0.71	1.15	1.32

L - LaneID; Ft - Fitting time; KSs - KS statistics; KSp - KS p value; ADs - AD statistics; ADp - AD p value; mFH - mean followers headway; sFH - std. followers headway

Appendix C

More Result Figures of Chapter 5

C.1 Using Pre-determined λ

Estimatied Parameters for pre-determined λ

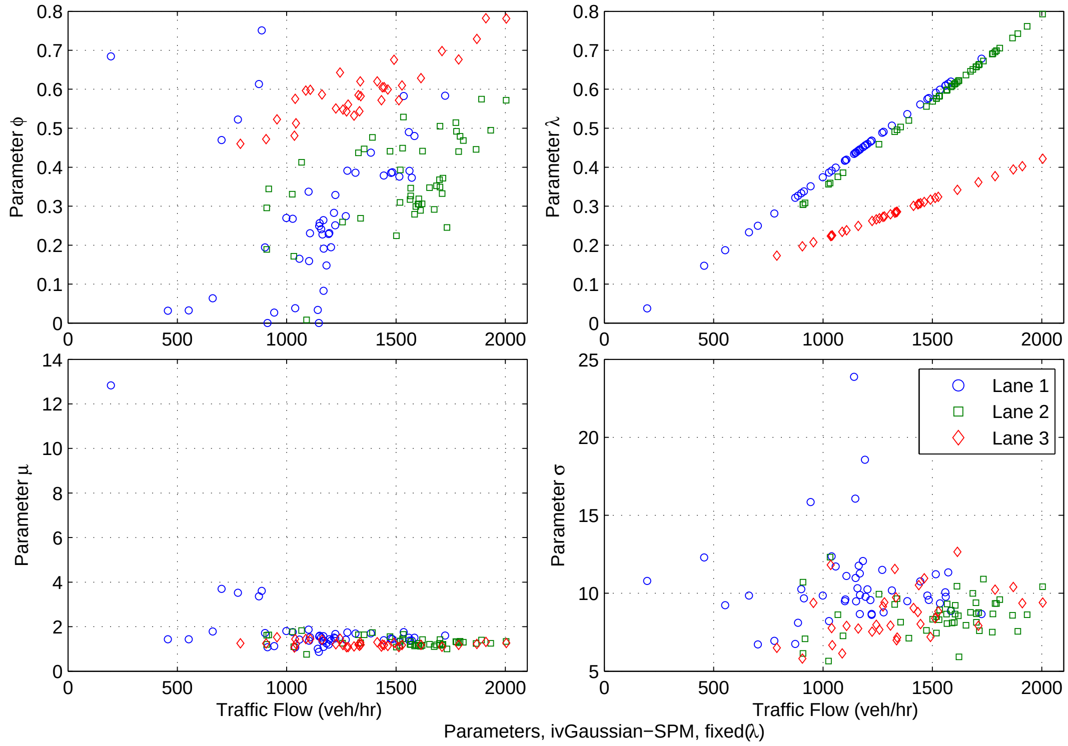

FIGURE C.1: Estimated parameters of IVG-SPM with pre-determined λ

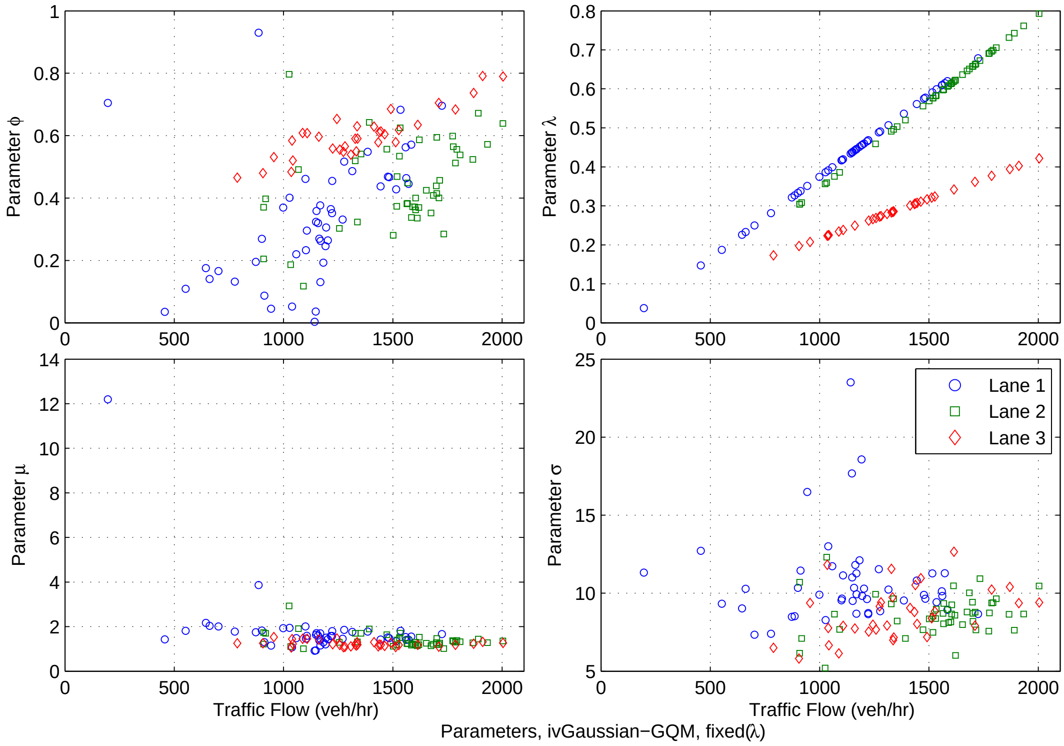

FIGURE C.2: Estimated parameters of IVG-GQM with pre-determined λ

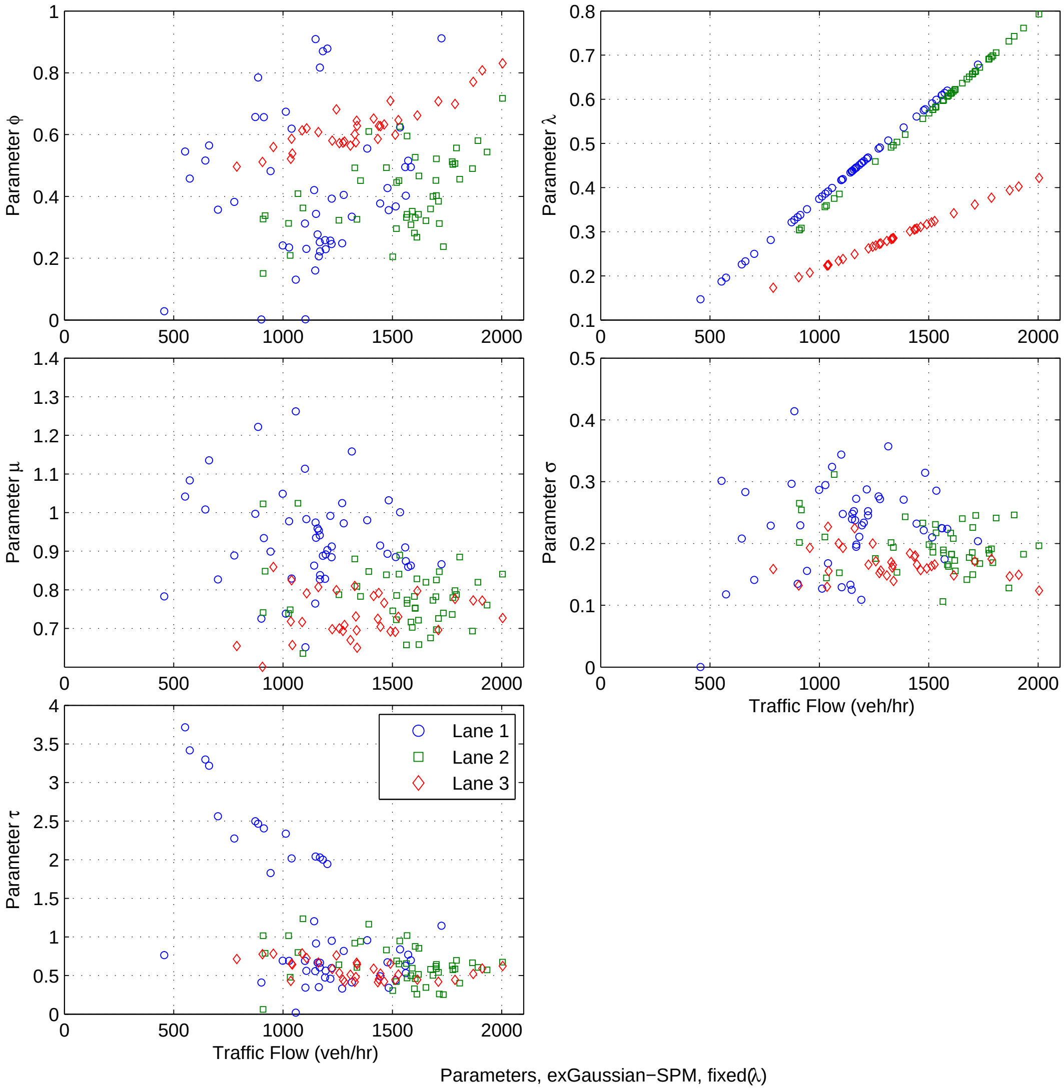

FIGURE C.3: Estimated parameters of EMG-SPM with pre-determined λ

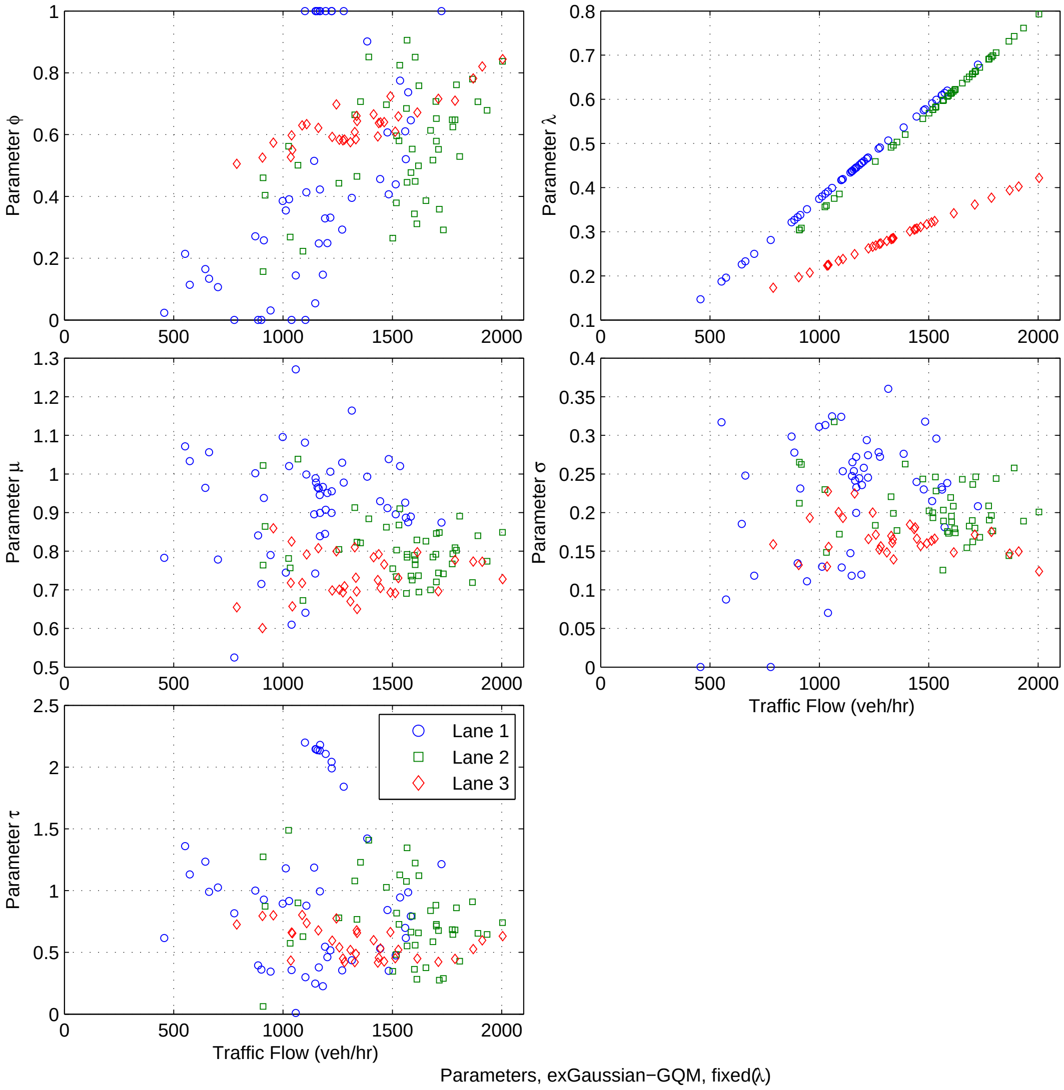

FIGURE C.4: Estimated parameters of EMG-GQM with pre-determined λ

C.2 Using Pre-determined ϕ, λ, σ (, τ)

Pre-determined ϕ, λ and σ for IVG-mixed models

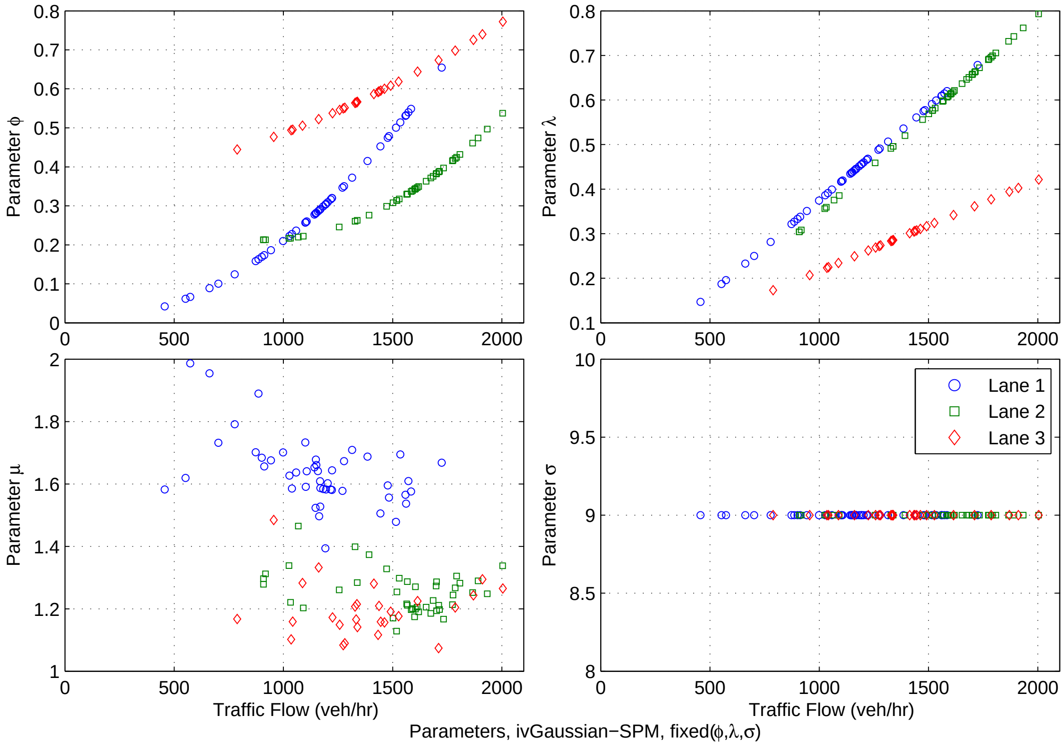

FIGURE C.5: Estimated parameters of IVG-SPM, fixed(ϕ,λ and σ)

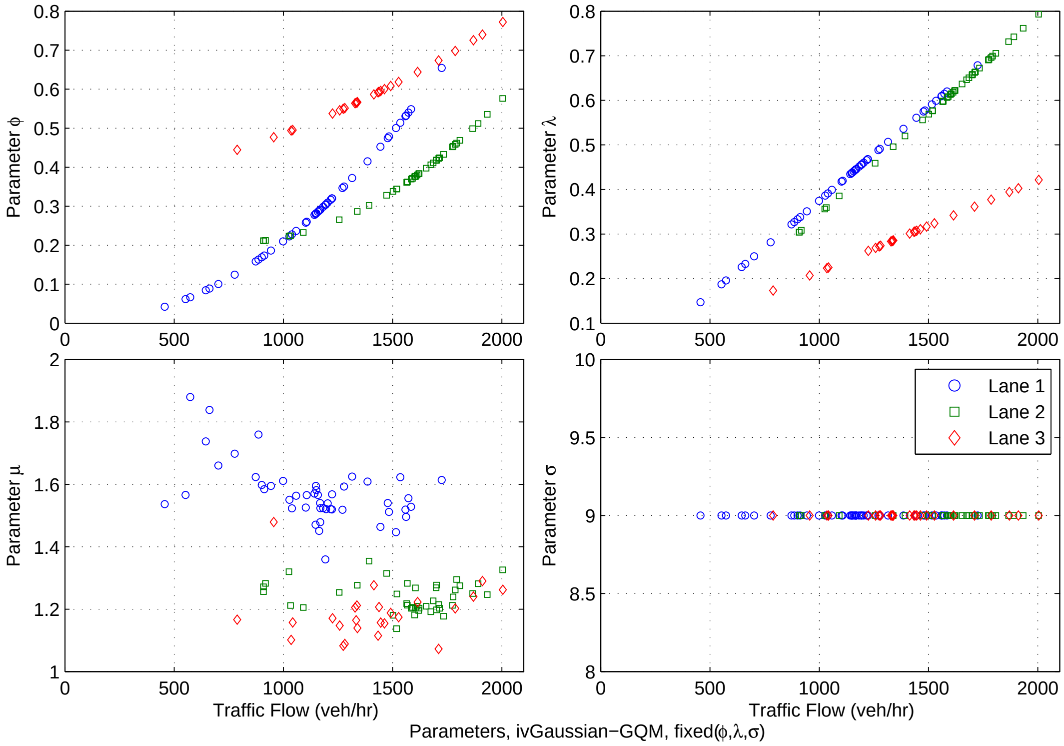

FIGURE C.6: Estimated parameters of IVG-GQM, fixed(ϕ,λ and σ)

Pre-determined ϕ, λ, σ and τ for EMG-mixed models

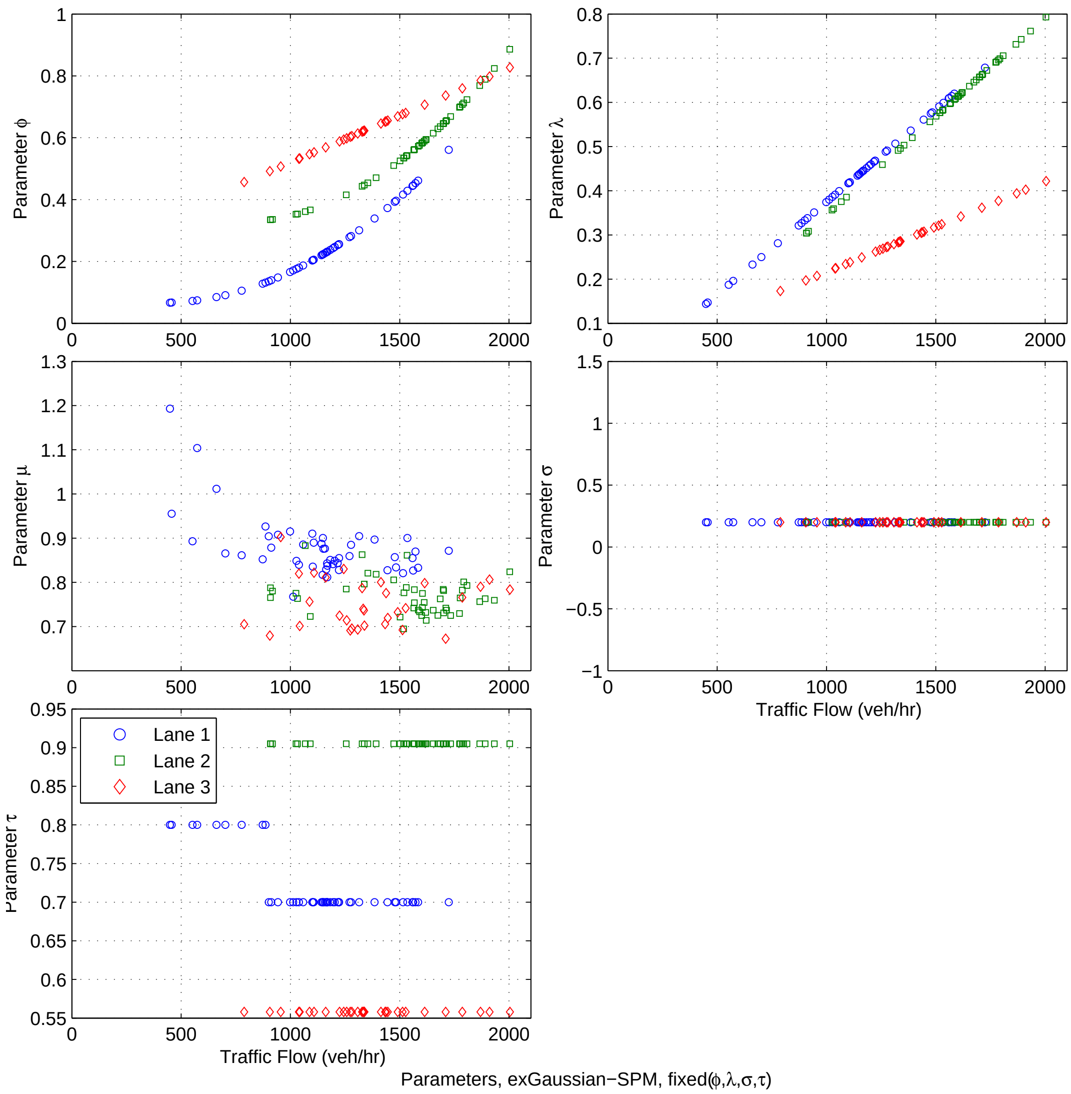

FIGURE C.7: Estimated parameters of EMG-SPM, fixed(ϕ, λ, σ and τ)

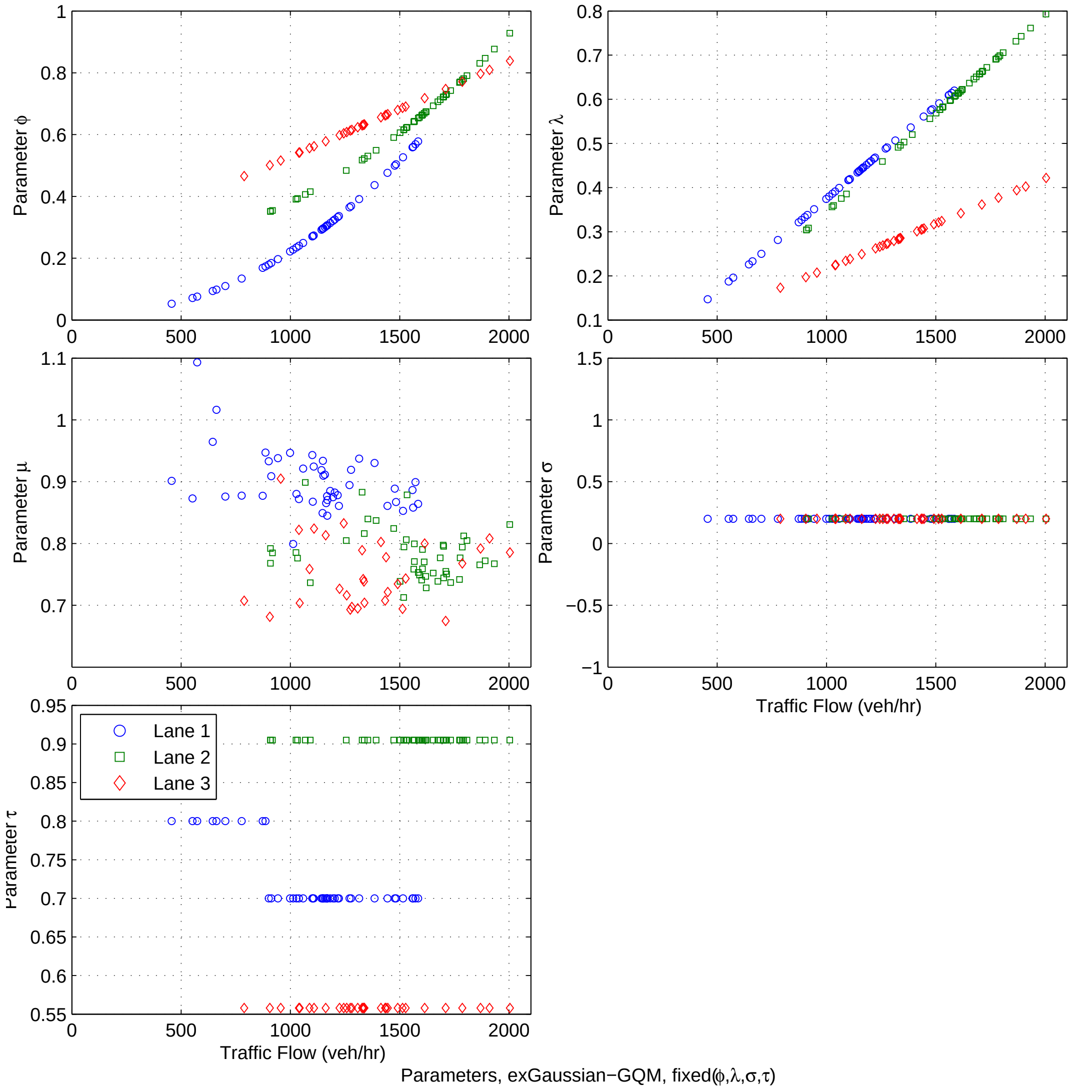

Figure C.8: Estimated parameters of EMG-GQM, fixed(ϕ, λ, σ and τ)

References

Adams, W. F. (1936). Road Traffic Considered as a Random Series. *Journal of the ICE*, 4(1):121–130.

Ashton, W. D. (1971). Distributions for Gaps in Road Traffic. *IMA Journal of Applied Mathematics*, 7(1):37–46.

Asmussen, S., Jensen, J. L., and Rojas-Nandayapa, L. (2014). On the Laplace transform of the lognormal distribution. *Submitted*.

Baayen, R. H. and Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, 3(2):12–28.

Blunden, W., Clissold, C., and Fisher, R. (1962). Distribution of acceptance gaps for crossing and turning manoeuvres. *Australian Road Research Board Proc*, 1(1):188–205.

Bock, R. K. and Krischer, W. (1998). *The Data Analysis BriefBook*. Springer Science & Business Media, Berlin/ Heidelberg.

Bolstad, B. M. (2004). *Low-level analysis of high-density oligonucleotide array data: background, normalization and summarization*. PhD thesis, University of California.

Bownan, K. O. and Shenton, L. R. (1982). Properties of estimators for the gamma distribution. *Communications in Statistics - Simulation and Computation*, 11(4):377–519.

Branston, D. (1976). Models of Single Lane Time Headway Distributions. *Transportation Science*, 10(2):125–148.

Branston, D. (1979). A Method of Estimating the Free Speed Distribution for a Road. *Transportation Science*, 13(2):130–145.

Breiman, L. and Lawrence, R. (1977). A simple stochastic model for freeway traffic. *Transportation Research*, II(1976):177–182.

Breiman, L. and Lawrence, R. L. (1973). Time scales, fluctuations and constant flow periods in uni-directional traffic. *Transportation Research*, 7:77–105.

Buckley, D. (1962). Road traffic headway distributions. In *Australian Road Research Board*, volume 1, pages 153–187.

Buckley, D. (1968). A semi-poisson model of traffic flow. *Transportation Science*, 2(2):107–133.

Bureau of Public Roads (1950). *Highway Capacity Manual*. U.S. Department of Commerce, Washington, D.C.

Chhikara, R. (1988). *The Inverse Gaussian Distribution: Theory: Methodology, and Applications*. CRC Press.

Chowdhury, M. A. and Sadek, A. W. (2003). *Fundamentals of intelligent transportation systems planning*. Artech House.

Cohen, A. C. and Whitten, B. J. (1980). Estimation in the Three-Parameter Lognormal Distribution. *Journal of the American Statistical Association*, 75(370):399–404.

Cohen, A. C. and Whitten, B. J. (1988). *Parameter Estimation in Reliability and Life Span Models*. M. Dekker, New York.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). On the LambertW function. *Advances in Computational Mathematics*, 5(1):329–359.

Cowan, R. (1975). Useful headway models. *Transportation Research*, 9(6):371–375.

Cox, D. R. and Lewis, P. A. W. (1966). *The statistical analysis of series of events*. Methuen & CO Ltd, London.

Cox, D. R. and Stuart, A. (1955). Some Quick Sign Tests for Trend in Location and Dispersion. *Biometrika*, 42:80–95.

Daou, A. (1964). The Distribution of Headways in a Platoon. *Operations Research*, 40(2):265–270.

Daou, A. (1966). On flow within platoons. *Australian Road Research*, 2(7):4–13.

Dawson, R. and Chimini, L. (1968). The hyperlang probability distribution-a generalized traffic headway model. *Highway Research Record*, (230):1–14.

Drew, D. R. (1967). Gap acceptance characteristics for ramp-freeway surveillance and control and discussion. *45th Annual Meeting of the Highway Research Board*, (157):108–143.

Drew, D. R. (1968). *Traffic Flow Theory and Control*. McGraw-Hill Book Company, New York.

Dunne, M., Rothery, R., and Potts, R. (1968). A Discrete Markov Model of Vehicular Traffic. *Transportation Science*, 2(3):233–251.

Figueiredo, L., Jesus, I., Machado, J. A. T., Ferreira, J., and de Carvalho, J. L. M. (2001). Towards the development of intelligent transportation systems. In *Intelligent Transportation Systems*, volume 88, pages 1206–1211.

Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2011). *Statistical Distributions*. John Wiley & Sons, fourth edition.

Forbes, T. W. (1963). Human Factor Considerations in Traffic Flow Theory. *Highway Research Record*, (15):60–66.

Forbes, T. W. and Simpson, M. E. (1968). Driver-and-Vehiele Response in Freeway Deceleration Waves. *Transportation Science*, 2(1):70–104.

Forbes, T. W., Zagorski, H. J., Holshouser, E. L., and Deterline, W. A. (1958). Measurement of Driver Reactions to Tunnel Conditions. *Highway Research Board Proceedings*, 37:345–357.

Gerlough, D. L. and Huber, M. J. (1975). Traffic flow theory : a monograph. *Transportation Research Board Special Report*, (165).

Green, M. (2000). "How Long Does It Take to Stop?" Methodological Analysis of Driver Perception-Brake Times. *Transportation Human Factors*, 2(3):195–216.

Greenberg, I. (1966). The log-normal distribution of headways. *Australian Road Research*, 2(7):14–18.

Greenwood, J. A. and Durand, D. (1960). Aids for Fitting the Gamma Distribution by Maximum Likelihood. *Technometrics*, 2(1):55–65.

Griffiths, J. and Hunt, J. (1991). Vehicle headways in urban areas. *Traffic Engineering and Control*, 32(10):458–462.

Gross, D. and Harris, C. M. (1985). *Fundamentals of Queueing Theory*. John Wiley & Sons, Inc., New York, NY, USA, second edition.

Ha, D. H., Aron, M., and Cohen, S. (2011). Variations in Parameters of Time Headway Models According to Macroscopic Variables, Fundamental Diagram, and Exogenous Effects. *Transportation Research Record: Journal of the Transportation Research Board*, 2260:102–112.

Ha, D. H., Aron, M., and Cohen, S. (2012). Time Headway Variable and Probabilistic Modeling. *Transportation Research Part C: Emerging Technologies*, 25:181–201.

Haney, S. (2011). *Practical applications and properties of the Exponentially Modified Gaussian (EMG) distribution*. PhD thesis, Philadelphia, Pa., Drexel University.

Heathcote, A. (2004). Fitting wald and ex-Wald distributions to response time data: an example using functions for the S-PLUS package. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc*, 36(4):678–94.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4):598.

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of experimental psychology*, 69(4):382.

Hoogendoorn, S. (2005). Unified approach to estimating free speed distributions. *Transportation Research Part B: Methodological*, 39(8):709–727.

Hoogendoorn, S. and Botma, H. (1996). On the estimation of parameters in headway distributions. *Proceedings of the 2nd TRAIL Ph.D. Congress.*

Hoogendoorn, S. and Botma, H. (1997). Modeling and Estimation of Headway Distributions. *Transportation Research Record*, 1591(1):14–22.

Hoogendoorn, S. and Bovy, P. (1998). New Estimation Technique for Vehicle-Type-Specific Headway Distributions. *Transportation Research Record*, 1646(1):18–28.

Johansson, G. and Rumar, K. (1971). Drivers' Brake Reaction Times. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 13(1):23–27.

Johnson, N. L., Kotz, S., and N., B. (1994). *Univariate discrete distributions*, volume 444. John Wiley & Sons, New York.

Jr., F. J. M. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78.

Kendall, M. G. and Ord, J. K. (1990). *Time Series.* Hodder Arnold, London, third edition.

Lacouture, Y. and Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, 4(1):35–45.

Lerner, N. D. (1993). Brake Perception-Reaction Times of Older and Younger Drivers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(2):206–210.

Lin, X. S. (1999). Laplace transform and barrier hitting time distribution. *Actuarial Research Clearing House*, 1:165–178.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford University Press.

Luttinen, R. (1996). *Statistical Analysis of Vehicle Time Headways.* PhD thesis, Helsinki University of Technology.

Matzke, D. and Wagenmakers, E. J. (2009). Psychological Interpretation of the Ex-Gaussian and Shifted Wald Parameters: a Diffusion Model Analysis. *Psychonomic bulletin & review*, 16(5):798–817.

May, A. D. (1961). Traffic Characteristics and Phenomena on High Density Controlled Access Facilities. *Traffic Engineering, Inst Traffic Engr*, 31(6):15–18.

May, A. D. (1965). Gap Availability Studies. *Highway Research Board Record*, (72):101–136.

May, A. D. (1990). *Traffic Flow Fundamentals.* Prentice-Hall, Inc, Upper Saddle River, NJ 07632 USA.

McDowell, M., Wennell, J., Storr, P., and Darzentas, J. (1983). Gap acceptance and traffic conflict simulation as a measure of risk. Technical report, Transport and Road Research Laboratory, Crowthorne, UK.

McGee, M. and Chen, Z. (2006). Parameter Estimation for the Exponential-Normal Convolution Model for Background Correction of Affymetrix GeneChip Data. *Statistical Applications in Genetics and Molecular Biology*, 5(1).

McKillup, S. (2005). *Statistics Explained: An Introductory Guide for Life Scientists.* Cambridge University Press, Cambridge, United Kingdom, 1st edition.

Miller, A. (1961). A Queueing Model for Road Traffic Flow. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(1):64–90.

Minderhoud, M., Botma, H., and Bovy, P. (1997). Assessment of Roadway Capacity Estimation Methods. *Transportation Research Record*, 1572(1):59–67.

NIST, Croarkin, C., Tobias, P., and Zey, C. (2001). *Engineering statistics handbook*. National Institute of Standards and Technology, US.

Ovuworie, G., Darzentas, J., and McDowell, M. (1980). Free Movers, Followers and Others: A Reconsideration of Headway Distribution. *Traffic engineering & control*, 21(8/9):425–428.

Ratcliff, R. and McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural computation*, 20(4):873–922.

Ratcliff, R. and Murdock, B. B. (1976). Retrieval Processes in Recognition Memory. *Psychological Review*, 83(3):190.

Schlotzhauer, S. D. (2007). *Elementary Statistics Using JMP (SAS Press)*. SAS Institute, Cary, NC, pap/cdr edition.

Silver, J. D., Ritchie, M. E., and Smyth, G. K. (2009). Microarray Background Correction: Maximum Likelihood Estimation for the Normal-Exponential Convolution. *Biostatistics (Oxford, England)*, 10(2):352–63.

Sivak, M., Post, D. V., Olson, P. L., and Donohue, R. J. (1981). Automobile Rear Lights: Effects of the Number, Mounting Height, and Lateral Position on Reaction Times of Following Drivers. *Perceptual and Motor Skills*, 52(3):795–802.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, 69(347):730–737.

Stuart, A. and Ord, J. K. (1991). *Kendall's Advanced Theory of Statistics Volume 2 5ed Classical Inference and Relationship*. Hodder General Publishing Division.

Sullivan, D. P. and Troutbeck, R. J. (1994). The use of Cowan's M3 headway distribution for modelling urban traffic flow. *Traffic engineering & control*, 35(7-8):445–450.

Summala, H. (2000). Brake Reaction Times and Driver Behavior Analysis. *Transportation Human Factors*, 2(3):217–226.

Tolle, J. E. (1971). The Lognormal Headway Distribution Model. *Traffic Engineering & Control*, 13(1):22–24.

Townsend, J. T. and Ashby, F. G. (1983). *Stochastic Modeling of Elementary Psychological Processes*. CUP Archive.

Transportation Research Board (1985). Highway Capacity Manual. In *TRB Special Report 209*. National Research Council, Washington, D.C.

Ulrich, R. and Miller, J. (1993). Information Processing Models Generating Lognormally Distributed Reaction Times. *Journal of Mathematical Psychology*, 37(4):513–525.

Van Zandt, T. (2000). How to Fit a Response Time Distribution. *Psychonomic bulletin & review*, 7(3):424–65.

Wasielewski, P. (1974). An Integral Equation for the Semi-Poisson Headway Distribution Model. *Transportation Science*, 8(3):237–247.

Wasielewski, P. (1979). Car-following Headways on Freeways Interpreted by the Semi-Poisson Headway Distribution Model. *Transportation Science*, 13(1):36–55.

Wilson, R. E. (2008). From inductance loops to vehicle trajectories. In *Proceedings of Symposium on the Fundamental Diagram*, volume 75, pages 134–143.

Wohl, M. and Martin, B. V. (1967). *Traffic System Analysis for Engineers and Planners*. McGraw-Hill Book Company, New York.

Young, M. S. and Stanton, N. A. (2007). Back to the Future: Brake Reaction Times for Manual and Automated Vehicles. *Ergonomics*, 50(1):46–58.

Zhang, G., Wang, Y., Wei, H., and Chen, Y. (2007). Examining Headway Distribution Models with Urban Freeway Loop Event Data. *Transportation Research Record*, 1999(1):141–149.